

Life as we know it

To understand the human genome, researchers must spread their wings to all branches of life.

As the technology for gene sequencing becomes more powerful, the prospect of personalized medicine draws ever closer. On 4 September, genomics pioneer Craig Venter followed in the footsteps of James Watson and published his complete genome sequence (see page 6). Plans are already being laid for hundreds more personal genomes to be sequenced (see *Nature* **447**, 358–359; 2007).

But a publication that will attract rather less attention than these personal genomes illustrates why researchers should ensure that genomic data are collected not just from humans and their closest relatives, but from every far-flung branch of the tree of life.

The paper in question focuses on segments of ‘ultraconserved’ DNA — sections that have stayed exactly the same throughout recent vertebrate evolution, and are identical in humans, rats and mice (see page 10). The available evidence suggests that this extreme example of DNA conservation is no accident: the sequence stays because there is a strong selective force weeding out mutations in it. In other words, it is likely to be important to its host.

Yet when researchers based at Lawrence Berkeley National Laboratory in California removed four pieces of ultraconserved DNA from different mice, it had absolutely no effect on the rodents (N. Ahituv *et al.* *PLoS Biol.* **5**, e234; 2007). This counterintuitive result contradicts predictions based on genetic conservation and the shaping of our genomes during evolution. Reconciling it with what scientists currently know would be easier if geneticists could figure out where ultraconserved DNA comes from, or what its function might be. Researchers have suggested that it could be involved in splicing RNA transcripts (J. Z. Ni *et al.* *Genes Dev.* **21**, 708–718; 2007) or in enhancing transcription (L. A. Pennacchio *et al.* *Nature* **444**, 499–502; 2006).

But so far there has been just one report on the origin of some ultraconserved DNA (G. Bejerano *et al.* *Nature* **441**, 87–90; 2006). A team from the University of California, Santa Cruz, traced the origins of one ultraconserved region back to a group of ancient fishes, including the coelacanth. This was only possible because another group of researchers had previously opted to sequence a few segments of coelacanth DNA. As Gill Bejerano, a former member of the Santa

Cruz team, says: “If the coelacanth people hadn’t been interested in that puny 1% of the genome we would not have the answer. Who knows what other information is out there?”

It is clear that efforts to understand the mechanisms of evolution will benefit from getting as much genetic information on as many diverse organisms as possible. As it happens, the US National Human Genome Research Institute (NHGRI) announced in May that it would add a fly and a worm to its ENCODE project, which aims to catalogue all the functional parts of the human genome, in order to meet that project’s human goals (see *Nature* **447**, 361; 2007).

The NHGRI and the other main public backer of genomics research in the United States, the Department of Energy, are each committed to comparative genomics. But under the influence of the ‘roadmap’ of the US National Institutes of Health (NIH), which emphasizes the translation of research findings into the clinic, the NHGRI is moving more forcefully into purely human genomics. The biggest new projects recently announced by the NHGRI are all human-centric, such as the Cancer Genome Atlas and a pair of initiatives to hunt for the genetic causes of human disease — the Genetic Association Information Network and the Genes, Environment and Health Initiative. The institute would also like to embark on a major human cohort study (F. S. Collins and T. A. Manolio *Nature* **445**, 259; 2007), and it is setting up a strategy for sequencing the genomes of our closest relatives, the non-human primates.

This is all understandable enough: the public is entitled to expect that the results of NIH research will be useful, when possible, to public health. But a better understanding of DNA function and the consequence of mutation will come only from generating more data from diverse genomes, backed by the bioinformatics capability that is needed to annotate them. That way scientists can learn more about the extent of DNA conservation throughout the living world — and, ultimately, tease out a deeper comprehension of the human genome. ■

“A better understanding of DNA function will come only from generating data from diverse genomes.”

The big splash

An unforeseeable chain of insights into an event 65 million years ago merits celebration.

When it comes to sensational science, the story of the asteroid impact some 65 million years ago at the boundary of the Cretaceous and Tertiary periods is hard to beat. The event itself must have been spectacular, with a vast fiery hole blown in Earth’s crust and atmosphere, and tsunamis racing out from the

impact point, kilometres tall. Its legacy, too, was impressive, dealing an apparent *coup de grâce* to a group of animals, the dinosaurs, that had dominated the continents for more than 100 million years. The science that brought this extraordinary event to light came with drama of its own — a startling announcement from a team that included a high-profile Nobel prizewinning physicist, Luis Alvarez, followed by decades of sometimes acrimonious debate as asteroid proponents and volcano supporters battled like titans.

The science of the K/T impact (K is the customary abbreviation for Cretaceous) began in a more modest way, with attempts to get a sense of how quickly a thin layer of clays in the Italian Apennines had

been deposited. No one foresaw that it would change how scientists and others see the world, and reintroduce catastrophism to the Earth sciences. Explanations that ignore the once-canonical principles of uniformitarianism — the gradualist paradigm in which the present is the key to the past — are now rife in studies of the history of Earth. A sense of Earth's abiding connection to the cosmos beyond is now more widely felt in the scientific community.

More broadly, the idea that Earth might be subject to such insult again at some time in the future has become a topic for box-office blockbusters and for sensible research and policy-making. The various surveys that were set in motion in the 1990s have reduced by almost an order of magnitude the risk of collision with a previously undiscovered asteroid that could drastically affect the global environment.

Against the background of these sweeping ideas, it is good to pause and take stock of the fact that there is still room for detail — indeed, for details that would have been scarcely imaginable when the first paper on the K/T impact was published in 1980 (L. W. Alvarez *et al.* *Science* **208**, 1095–1108; 1980). Geologists are now seriously proposing that they might be able to date the events of 65 million years

ago to within 25,000 years — which is to say by better than 1 part in 2,500 (see page 20). Meanwhile, astronomers are saying with about 90% confidence that the asteroid that struck Earth on that darkest of days was a sibling of the asteroid Baptistina. Both seem to have been chips off a larger block that was destroyed 160 million years ago in a catastrophe of its own (see pages 30 and 48).

When the idea of an impact was first mooted, the notion of fleshing out the story to this degree of elaboration — with the impact site characterized, the timing set to extraordinary accuracy, and even the source of the utterly destroyed impactor narrowed down to a particular parent — was all but unthinkable.

The ability of diverse sciences to collectively provide a coherent story about a time so distant is the sort of triumph that enthusiasts for interdisciplinary approaches should single out for praise and emulation. They should also celebrate the scientific pursuit of ever finer detail. After all, this Earth-shattering narrative began with just such attention to a centimetre-thin stratum. ■

"It is good to pause and take stock of the fact that there is still room for detail."

A question of breeding

China needs to rethink its approach to conservation if it wants to protect its endangered tigers.

Hengdaohezi in China's Heilongjiang province is home to the largest tiger-breeding facility in the world. Portraying itself as a centre for conservation research, it says its primary goal is to breed and save the endangered Siberian tiger. Unfortunately, the centre seems mainly to be a spectacle for tourists. And there are disturbing signs that it may soon begin to defy international conservation rules, by farming tiger body parts for medicinal use (see page 16).

The facility is breeding at least 100 tigers a year but, by taking no note of which animals breed with each other, it may be diluting the genetic stocks of the captive cats indiscriminately. Staff at the centre seem unclear on how to promote conservation, stating a desire, for example, to reintroduce the tigers into the wild without planning the measures that would be needed to enable them to survive.

Reliable information about the facility's plans for its 700 animals is hard to come by. Park officials have denied broad proclamations by government officials of plans to release some 600 tigers into the wild ahead of next year's Beijing Olympics. If this plan went ahead, it would almost certainly fail, exposing animals that have grown accustomed to human attention to the mercy of their wild counterparts, possibly with fatal results.

The pattern at the Hengdaohezi facility is, unfortunately, typical of animal conservation efforts in China. Despite recent improvements in its breeding strategy, the Wolong Giant Panda Research Center in Sichuan province, for example, has not been sufficiently open about incidents such as the death of Xiang Xiang, a panda released into the wild last year.

The outlook for the tigers at Hengdaohezi is also coloured by officials at the State Forestry Administration and elsewhere saying that China plans to reopen trade in tiger body parts for use in traditional Chinese medicine. Indeed, farming the tigers for this use was the initial purpose of the Hengdaohezi centre when the government opened it in 1986. It is only thanks to a ban enacted in 1993 in compliance with the Convention on International Trade in Endangered Species of Wild Fauna and Flora that the facility switched to tourism and conservation research.

It is easy, of course, for outsiders to criticize such practices without acknowledging that nations have more things to think about than animal conservation. And some observers have voiced support for the proposed trade in tiger parts, saying that it could allow local people to make a living, while potentially relieving poaching pressure on the wild tiger population.

But most conservationists have condemned any repeal of the ban on the tiger trade. Even if, as proponents suggest, a legal trade did reduce the price of body parts, prices would remain high enough to encourage poaching. Conferring legal status on the trade would, in the end, increase trading volumes and put more pressure on the wild population.

At the Hengdaohezi conservation facility, meanwhile, poor breeding management and detachment from the international conservation community has produced a group of docile animals that would stand little chance of survival in the wild.

Breeding and conservation are two very different things. If China is serious about conserving Siberian tigers, or other endangered species, it should concentrate more on the conservation of natural habitats, together with the careful monitoring of animal health, and less on the rapid breeding of animals in confinement. ■

"Giving legal status to the trade in tiger body parts would put more pressure on the wild population."

RESEARCH HIGHLIGHTS

Double fault

Geology **35**, 855–858 (2007)

Earthquakes in two separate fault systems in Southern California seem to be linked, with increased seismic activity in one zone being matched by a lull in the other.

James Dolan at the University of Southern California in Los Angeles and his colleagues say their data suggest that the release of strain in tectonic plates isn't entirely random, but is controlled by long-distance and long-term fault interactions. The researchers compared records of seismicity in a complex fault system under Los Angeles with records from a region to the northeast, the eastern California shear zone in the Mojave Desert (pictured).

The team is now drilling cores, checking faults and charting earth movement with global positioning systems to test the theory.



J. DOLAN

CLIMATE SCIENCE

Pouring over the past

Geophys. Res. Lett. **34**, L16707 (2007)

A 27-centimetre long stalagmite from a cave in central India has provided a 900-year record of rainfall from India's summer monsoon. The data span from 600 to 1500 ad, before instrumental records began.

The stalagmite, dated by analysis of uranium-series isotopes, suggests periods of substantially reduced monsoon rainfall that square with historical accounts of droughts and famines. Monsoon failures of this scale have not occurred during the 150-year instrumental period. The data pick out some of the same climate-related changes in the monsoon seen in records from marine sediments, stalagmites and pollen collected from elsewhere. Ashish Sinha of California State University, Dominguez Hills, in Carson and his colleagues suggest that differences between the records on shorter timescales are due to uncertainties in dating.

NANOTECHNOLOGY

Spare the rod

Nano. Lett. doi:10.1021/nl071615y (2007)

Rod-shaped nanoparticles can be sorted from a jumble of particles such as that pictured (right) with a technique borrowed from molecular biology, report researchers from the University of Mainz in Germany.

Carsten Sönnichsen and his colleagues found that gel electrophoresis, commonly used to separate DNA or proteins, also works on mixtures of silver nanoparticles coated with a charged polymer. On the whole, rod-shaped particles move more slowly through the gel than round or triangular particles.

Current synthesis techniques for silver nanoparticles typically produce all shapes at once. Filtering out the rods is desirable because they have applications in imaging and sensing. The researchers are now working to improve and scale up the technique.

NEUROSCIENCE

One and the same

Science **317**, 1230–1233 (2007)

Researchers have shown that the learning and retrieval of a memory in mice have at least some neurons in common.

Mark Mayford and his colleagues at the Scripps Research Institute in La Jolla, California, created a transgenic mouse in which it is possible to tag and identify neurons that are active at different times. Mice engineered in this way were tested to identify neurons involved first in learning, and then later in retrieval, of a conditioned fear response. In the basolateral amygdala, a brain

region important for emotional memory, some neurons were involved in both processes. Moreover, the number of reactivated neurons correlated with the strength of the contextual fear memory retrieval.

ASTRONOMY

Brief encounters

Astrophys. J. **666**, 346–360 (2007)

A survey of more than two decades of radio-astronomy data has turned up a handful of mysterious objects, most of which appeared in only one of the once-weekly observations.

Goffrey Bower and his colleagues at the University of California, Berkeley, reviewed 22 years' worth of data from the Very Large Array near Socorro, New Mexico, spotting ten transient objects. These had some features reminiscent of the afterglows of gamma-ray bursts, but not all the details matched. The group believes that some of the objects may have been remnants of a previously unknown type of stellar explosion, or a rare, very powerful type of eruption from a faint, low-mass star. Bower says that observations are now being planned to look for such events.

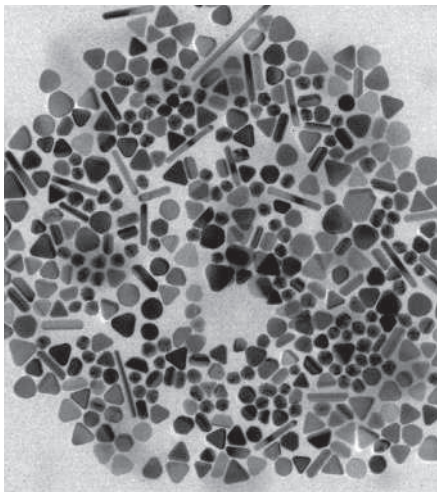
GENETICS

Tall story

Nature Genet. doi:10.1038/ng2121 (2007)

A genetic study involving more than 34,000 people has identified a single base change in a gene known as *HMGA2* that boosts overall body height. This is thought to be the first genetic variant decisively correlated with normal height variation — a trait potentially influenced by hundreds of genes.

The researchers, led by Timothy Frayling at the Peninsula Medical School in Exeter,



UK, report that people with two copies of the 'tall' variant are on average almost a centimetre taller than those with a double copy of the 'short' version. The precise function of *HMGA2* is not yet clear, although it is involved in unravelling the chromatin in which chromosomes are packaged, suggesting that it may influence cell growth and division.

CHEMISTRY

A simple solution

Science **317**, 1189–1192 (2007)

Our understanding of how molecules will react is usually gleaned from experiments carried out with organic solvents available in standard chemistry labs. For natural products that come from water-living organisms, it might help to think about things differently.

Timothy Jamison and Ivan Vilotijevic at Massachusetts Institute of Technology in Cambridge have synthesized the core piece of a famously hard-to-make marine molecule by working in neutral water. The molecule, which causes the toxicity associated with 'red tide' algal blooms, has a ladder-like arrangement of rings. It was thought that this structure might assemble through a cascade of reactions, but not until Jamison and Vilotijevic tried it in water did the molecule zip together with ease.

QUANTUM PHYSICS

Up in the air

New J. Phys. **9**, 254 (2007)

Although the Casimir force is generally regarded as attractive, theory predicts that this force, which acts between two closely spaced surfaces because of quantum fluctuations, can be made repulsive. This effect could be useful in nanoengineering (see *Nature* **447**, 772–774; 2007). Ulf Leonhardt and Thomas Philbin

of the University of St Andrews in Scotland suggest one way to turn the pull into a push.

They argue that a 'left-handed metamaterial' — comprising an array of electrical and magnetic components that bend light the 'wrong' way — placed between two mirrors may make the Casimir force repulsive. The researchers estimate that the repulsion should be strong enough, with metamaterials within experimental reach, to levitate a 0.5-micrometre thick aluminium foil in a vacuum.

IMMUNOLOGY

Itchy and scratchy

Nature Immunol. doi:10.1038/ni1503 (2007)

Mast cells, a type of proinflammatory immune cell, have been unjustly accused of exacerbating the blistering itch of poison ivy (pictured below) and sunburn, say researchers.



Mast cells have been shown to increase short-term swelling in response to skin irritation. But Stephen Galli and his co-workers at Stanford University School of Medicine in California have found that instead of making the itch worse, mast cells reduce long-term inflammation by secreting a protein known as interleukin-10. Furthermore, chemical irritants elicited a more severe response in mice engineered either to lack mast cells, or to lack interleukin-10, than they did in normal mice. The results could spur development of novel anti-inflammatory therapies.

METHODS

Caught in the act

Nature Biotechnol. doi:10.1038/nbt1328 (2007)

A new proteomics technique can screen drugs for activity against hundreds of protein kinases at once. The technique has revealed two previously unidentified targets of imatinib, a cancer treatment that inhibits the BCR-ABL kinase and some other proteins.

Developed by Gerard Drewes and Bernhard Kuster at Cellzome in Heidelberg, Germany, and their colleagues, the approach relies on chemically coated beads that latch onto the hundreds of kinases, kinase-bound proteins and related purine-binding proteins found in cells. The kinases and other proteins can't attach to the beads if they are bound to a drug, so comparison of beads exposed to the contents of normal cells with those exposed to drug-treated cells shows which molecules the drug has blocked.

The bead-bound proteins are detected by mass spectrometry. This can also measure whether the drug has induced phosphorylation of the enzymes, a chemical modification that affects their activity.

R. PLANCK/NHPA

JOURNAL CLUB

Drew Endy
Massachusetts Institute of
Technology, Cambridge, USA

A biological engineer searches for simplicity.

Several years ago, a good colleague suggested that I read about a discussion held in 1864 on nuts and bolts (*J. Franklin Inst.* **77**, 344–351; 1864). The focus was a paper by one William Sellers that argued for the adoption of a uniform system of screw threads — 60° angles, squared off along the edges.

Machinists across the United States eventually started producing nuts and bolts according to Sellers' scheme. As a result, hardware stores now offer a wide selection of standardized parts that can be used in combination and behave as expected.

Inspired by this example and others, I have been studying how synthetic biological parts might be made as regular and easy to use as Sellers' nuts and bolts.

The starting complexity of nature has led some distinguished researchers to doubt such work is practical. But given that there

has been little research on manufactured bio-simplicity, this seems premature.

And there are examples: a team at the California Institute of Technology in Pasadena recently developed a uniform system for engineering simple biological switches made from ribonucleic acids (M. N. Win and C. D. Smolke *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0703961104; 2007).

The 'nuts and bolts' of the switches are RNA sensor and actuator domains. The method for combining any sensor domain to

an actuator domain through a third communication domain provides the 'uniform screw threads'. Because such switches are produced by a standard process, many switches could be quickly programmed to control diverse cellular functions in response to myriad molecular inputs, from small molecules, to peptides, to nucleic acids.

I suspect that further efforts to engineer biological simplicity will have similarly powerful results.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS

All about Craig: the first 'full' genome sequence

Controversial genomics pioneer Craig Venter has sequenced his own genome. In a preliminary analysis published this week (S. Levy *et al. PLoS Biol.* 5, e254; 2007), Venter's team has picked apart the sequences belonging to both chromosomes in each of the 23 chromosome pairs found in his cells, providing the first glimpse of the variation found within a single genome.

The paper also highlights some features of Venter's genome. The sequence of his *ABCC11* gene, for example, indicates that Venter is likely to have wet earwax, as opposed to dry. In a finding likely to disappoint some critics of the maverick scientist, Venter has four repeat sequences located just before his *MAOA* gene. Having only three of the repeats is associated with an increased risk of antisocial behaviour. On a more sobering note, Venter's *APOE* sequence is associated with a higher risk of Alzheimer's and cardiovascular diseases, and his *SORL1* gene also contains several variants that are associated with Alzheimer's disease.

For more about his personal genomic and

family history, Venter aficionados can refer to his upcoming book, *A Life Decoded*. Venter has placed references to how his genome sequence may have affected his life throughout the book, says Jan Witkowski, executive director of the Banbury Center at Cold Spring Harbor Laboratory in New York, who is reviewing the book for the 4 October issue of *Nature*.

Venter's sequence also provides important new information about the human genome. Previous sequencing efforts have not distinguished between the two copies of each chromosome, or even between DNA from different donors. "What we were doing was mixing up alleles," says Samuel Levy — who led the study at the J. Craig Venter Institute in Rockville, Maryland — referring to DNA sequences from specific spots on the chromosome. "We were creating Frankenstein versions of chromosomes."

This time, armed with an additional 13 million sequences — added to the 19 million generated from Venter's own DNA during the first genome project — and fresh algorithms



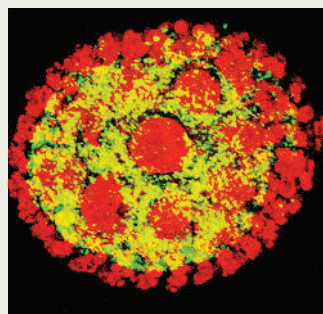
designed to pick apart sequences from different versions of the same chromosomes, Levy and his team could look at the variation within the genome. They found more than 4 million vari-

M. NAGLE/GETTY

Genomes within genomes

Another team of genome researchers at the J. Craig Venter Institute in Rockville, Maryland, which has been investigating the DNA of a rather less salubrious organism, this week reports a surprise discovery: the DNA of fruitfly *Drosophila ananassae* contains the entire genome of a parasitic bacterium of the *Wolbachia* genus. Smaller parts of the parasite's genetic material also turned up in worms and wasps.

Bacteria commonly swap DNA with each other. But transfer of bacterial genes into animals was thought to be rare. The new work, published in *Science* (J. C. Dunning Hotopp *et al. Science* doi:10.1126/science.1142490; 2007), suggests that gene flow from bacteria to animal hosts happens on a larger scale and more commonly than suspected. And it hints that the



Invader: *Wolbachia* bacteria (yellow) inside a developing fruitfly egg (red).

bacterial genome may have provided some sort of evolutionary advantage to its host. "You're talking about a significant portion of [the fruitfly] DNA that is now from *Wolbachia*," says Julie Dunning Hotopp, who led the study. "There has to be some sort of selection to carry around that much extra DNA."

But Dunning Hotopp's former

colleague Jonathan Eisen of the University of California, Davis, contests this. "One cannot conclude that some DNA is advantageous simply because it is there," he says.

Up to 75% of insect species are plagued by *Wolbachia*, which lives inside testes and ovaries and passes from one female generation to another through infected ova. To ensure its spread, *Wolbachia* can skew insect birth ratios towards females and even prevent infected males from successfully mating with disease-free females. The bacterium's close association with egg cells means there's ample opportunity for bacterial DNA to get permanently sewn into a host's nuclear genome, says Dunning Hotopp.

When Dunning Hotopp and her colleagues analysed the DNA of *D. ananassae* uninfected by

Wolbachia, they found 44 of the 45 *Wolbachia* genes they searched for. Because these selected genes are so widely spread throughout *Wolbachia* DNA, this suggests that the rest of its more than 1-million-base-pair genome is also likely to be found in fruitflies.

Many of the *Wolbachia* genes were infiltrated by strands of insect DNA that jump around the genome, and so are unlikely to be functional. But the researchers showed that at least 28 of the bacteria's 1,206 genes are active in the flies. More genes that have seeped from bacteria into animals are certain to be found, the researchers say, particularly in reptiles and amphibians. But finding bacterial genes in mammals is unlikely because no bacteria are known to infect their sperm and egg cells. ■

Ewen Callaway

SCIENCE



GENETICS NEWS

Find all our coverage of genetics research in one place.

www.nature.com/news/subject/genetics



The man within: enthusiasts will be able to examine the genetic make-up of Craig Venter.

ations between the two sequences, including single nucleotide differences, sequence insertions and deletions, and differences in the number of copies of a given gene. Some 44% of Venter's genes contained a genetic difference between copies found on each chromosome. Venter's two sets of chromosomes differed by 0.5%, suggesting that there may be seven times more DNA variation than previously expected, says Levy.

This approach provides a clearer picture of the human genome, says Edward Rubin, director of the Joint Genome Institute in Walnut Creek, California. Before, the sequence gave a "statistical view" of the genome, Rubin says. "And in fact the genome is not statistical, it's really a linear array of bases."

Venter notes that single genetic changes are unlikely to seal his fate. "I take it very seriously," he says. "But most diseases are going to be some huge compilation of human factors and environmental factors." Witkowski agrees, but says that reading about someone's genome can strike an emotional chord. "Somehow there's a sense that when you tell people that sequence, you're telling them in a very deep way about yourself," he says. "It's like looking at their medical records."

Heidi Ledford

See Editorial, page 1.

DNA probe finds hints of human

A groundbreaking analysis of Neanderthal DNA that suggested they interbred with humans was based on samples contaminated with human DNA, a new study suggests.

The study¹, published on 28 August in *PLoS Genetics*, reanalysed about one million base pairs of fossilized Neanderthal DNA that had been analysed in a paper published last November in *Nature*². The *Nature* paper and a paper in *Science*³ published the same week on 65,000 base pairs were the first reports on Neanderthal nuclear DNA.

But around 80% of the sequences in the *Nature* paper are modern human DNA, not Neanderthal, claims Jeffrey Wall, an evolutionary geneticist at the University of California, San Francisco, who led the *PLoS Genetics* study. This indicates that human genetic material was somehow introduced into the samples. This known risk is increased by the closeness of the two species — the 3-billion-base-pair genomes of a human and a Neanderthal differ by less than 0.5%.

The results in the *Nature* paper suggested that there was interbreeding among Neanderthals and humans in their common European home ground before Neanderthals became extinct 30,000 years ago. The *Science* article found no genetic evidence of interbreeding.

Svante Pääbo, senior author of the *Nature* paper, concedes that his group at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, had problems with contamination. These

prompted him to change laboratory procedures and to add controls late in 2006, after the paper was published. "I agree with [Wall's] analysis," Pääbo says. "Their observations are formally correct."

Pääbo's co-author Michael Egholm, who is research vice-president at 454 Life Sciences in Branford, Connecticut, adds: "There is no denying contamination. It was one of the dangers of doing this." But ongoing analysis indicates that human contamination in their study was just 30%, Egholm adds.

There had been intense debate over the contrasting results in the *Nature* and *Science* papers, which analysed the same 38,000-year-old Neanderthal bones from Croatia using different sequencing methods. Pääbo's group used 454's rapid 'direct sequencing' approach, whereas the *Science* team, led by Edward Rubin of the Joint Genome Institute in Walnut Creek, California, used a traditional method using cloned DNA and bacteria to generate the base pairs.

The studies gave different estimates for the time Neanderthals diverged from humans — the *Science* article pegged it at 706,000 years ago, whereas the *Nature* paper set it at 516,000 years ago. Wall's study confirms the 706,000-year divergence date. The probable human DNA contamination led to the more recent date and may have led to the suggestion of later interbreeding, Wall says.

The *Nature* paper also found more similarities between genetic variations called SNPs (single nucleotide polymorphisms) in Neanderthal and human DNA than the *Science* paper, even after allowing for the *Nature* group's larger number of base pairs. Pääbo's team reported that about 30% of the SNPs in the Neanderthal DNA are derived — that is, the mutations occurred — in today's humans.

Pääbo acknowledges there is "a potential problem" with the presence of these human SNPs in the Neanderthal sequence. These same discrepancies were noted by Rubin's group. "We had concerns," says Rubin. "We suspected some of the issues raised by Wall."

Both Pääbo's and Rubin's groups expect to publish further Neanderthal sequences from other specimens that each group is studying. Pääbo and Egholm say their analysis will address the anomalies in their *Nature* paper. ■

Rex Dalton



Neanderthal DNA has been reanalysed, leading to suggestions of human DNA contamination.

1. Wall, J. D. & Kim, S. K. *PLoS Genet.* doi:10.1371/journal.pgen.0030175.eor (2007).
2. Green, R. E. et al. *Nature* **444**, 330–336 (2006).
3. Noonan, J. P. et al. *Science* **314**, 1113–1118 (2006).

J. TRUEBA/MSF/SPL

ON THE RECORD

“I was too black to be a scientist and too educated to be a lot of other things.”

Guyanese novelist E. R. Braithwaite explains why 1950s British society led him to become a schoolteacher instead of a researcher after gaining his Cambridge physics doctorate.

SCORECARD

**Hybrid engines**

The technology enjoyed perhaps its finest moment as the Toyota Supra HV-R claimed victory in a Japanese 24-hour endurance race, leaving non-hybrid rivals trailing in its wake.

**Brakes**

A study of urban traffic in Stockholm has shown that brake linings (unfortunately an essential component even of hybrid vehicles) give off a host of toxic metals, including copper, zinc and antimony.

ZOO NEWS

Elephant goes cold turkey

An elephant called Big Brother has become possibly the first to rid itself of heroin addiction with the aid of methadone. Illegal traders had spent two years feeding him bananas laced with heroin in an effort to control him and his herd, before officials in Yunnan, China, stepped in to put him back on the straight and narrow.

ROBOT NEWS

Shot in the arm

Engineers at Vanderbilt University have developed a robotic arm that could help amputees avoid the need to carry around bulky battery packs. The solution? The same type of rocket motor used to steer the space shuttle in orbit — which, as you might expect, packs quite a punch per pound.

Sources: BBC, Sci. Am., Environ. Sci. Technol., Reuters, ABC News

Turkish physicists face accusations of plagiarism

More than a dozen theoretical physicists at four universities in Turkey seem to be involved in a massive plagiarism scandal.

Almost 70 papers by 15 authors have been removed from the popular preprint server arXiv, where many physicists post their work, by the server's moderators. They allege that the papers plagiarize the works of others or contain inappropriate levels of overlap with earlier articles. This is probably the largest single incident of its sort ever seen on the server, according to physicist Paul Ginsparg of Cornell University in Ithaca, New York, and founder of arXiv. “What these guys did was way over the line,” he says.

Not all the researchers or universities allegedly involved in the scandal could be reached as *Nature* went to press, but the author with the most papers withdrawn called the accusations “ridiculous”. “We carried out a good collaboration,” says Mustafa Salti, a graduate student at the Middle East Technical University (METU) in Ankara whose name is on 40 of the withdrawn papers. “Most of our papers have been published in the science citation index journals. Until now no one has claimed that we plagiarize.”

The trouble began last November, when Salti and another graduate student, Oktay Aydogdu, underwent oral examinations for their PhDs. Although both had an extensive list of publications in gravitational physics, they struggled to answer even basic, high-school-level questions, according to Özgür Sarioğlu, an associate professor at METU. “They didn't know fundamental stuff like newtonian mechanics,” he says.

Suspicious, one of Sarioğlu's colleagues, Ayşe Karasu, began to look through the duo's publication record. Using Google, she quickly turned up a paper from which it seemed the students had lifted several lengthy sections. By mid-February, faculty members had identified dozens of articles on arXiv that they say seemed to be partly or completely plagiarized.

They notified Ginsparg, who began his own investigation. So far, the search has turned up 67 papers, about half of which have appeared in low-profile peer-review journals. In addition to the two students at METU, the papers were authored by professors and students at three other universities: Dicle University in Diyarbakir, the University of Mersin in İçel, and

Onsekiz Mart University in Çanakkale. Sarioğlu says that most of the plagiarism seems to have come from two rival groups, one at Onsekiz Mart and the other at Mersin and Dicle.

Many of the papers concern an obscure theory of gravity known as the Møller version of general relativity. Few people would be likely to check such work, allowing the students and professors to build their publication record without fear of being caught, says Ginsparg. “They were following the optimal strategy.”

“They're isolated, their English is bad, and they need to publish,” says Sarioğlu. “So they plagiarize, I guess,” he says of the alleged plagiarizers.

A recent analysis turned up numerous examples of plagiarism on the arXiv server (see *Nature* 444, 524–525; 2006). Ginsparg says that it's not uncommon for scientists with a poor command of English to plagiarize introductions or background paragraphs from earlier work, often adding an appropriate citation. He thinks that although such practices are ethically questionable, it is inappropriate to be overly draconian.

Katepalli Sreenivasan, director of the International Centre for Theoretical Physics in Trieste, Italy, which has a programme of collaboration with physicists from the developing world, agrees. “There are some cultures in which plagiarism is not even regarded as deplorable,” he says. Problems of academic integrity come up frequently at the centre, and are dealt with on a case-by-case basis, he adds.

But both think that the Turkish case clearly crosses the line. “It's dishonest and sloppy,” Ginsparg says. He adds that, although arXiv normally declines to publicize such incidents, the size and scope of this case made it an exception.

Salti and Aydogdu have been suspended for two terms, according to Sarioğlu. They cannot be expelled because METU's ethics policies for students do not include rules on plagiarism. But, he says, it is unlikely they will be able to continue at the university. Meanwhile, Salti says that he and several authors plan to bring a lawsuit against arXiv. It remains unclear whether the other three universities involved have taken any action.

Geoff Brumfiel

“There are some cultures in which plagiarism is not even regarded as deplorable.”

T. GAINEV/ALAMY



SNAPSHOT Rare beef

Africa's impressively horned Ankole cows are one of many threatened indigenous livestock breeds. A survey of cattle, pigs, goats, sheep and poultry in 169 countries found that native breeds are being supplanted by high-yielding commercial strains, which may

not be as resistant to disease and drought.

"Valuable breeds are disappearing at an alarming rate," says survey leader Carlos Seré, director-general of the International Livestock Research Institute in Nairobi. At a meeting this week of policymakers, breeders and livestock scientists, he called for the creation of gene banks to conserve sperm and eggs

from rare breeds.

The survey adds to fears that damaging levels of inbreeding exist even among the most abundant commercial breeds. For example, the world's most widespread cattle breed, Holstein-Friesian, is thought to be almost exclusively descended from just a few dozen prized individuals owned by breeding companies. ■

Michael Hopkin

Biotech crop rules get rewrite

A process aimed at revising the regulation of genetically modified organisms (GMOs) has been launched by the US Department of Agriculture (USDA). But critics fear the changes will not go far enough to protect the environment and public health.

The USDA is one of three US agencies responsible for regulating GMOs, along with the Food and Drug Administration and the Environmental Protection Agency. A draft environmental-impact statement released in July gives the first glimpses of how the USDA rules might change. It proposes to expand its authority from plants that might endanger other plants to "the full range of potential agricultural and environmental risks posed by these organisms, including risks to public health".

The agency is also suggesting a tiered programme, with different regulations for different levels of risk, and special rules for

organisms that produce compounds used in industry or pharmaceutical products. Other proposals include allowing regulation of a crop that has previously been officially "deregulated" and creating a new category for permitted "low level presence" of GMOs in food and seed crops.

Karen Perry Stillerman, a senior analyst for the Union of Concerned Scientists, is underwhelmed. "We would like to see a ban on growing pharmaceutical and industrial compounds in food crops outdoors," she says.

"This is the kind of feedback we are trying to get," says USDA spokeswoman Rachel Iadicco. Public comment will be collected until 11 September; then the USDA will probably write a proposed rule that would have its own public comment period. Only then will the final rule be published. "It is a lengthy process," says Iadicco. ■

Emma Marris

The dune chorus

The latest explanation of the mysterious 'song of the dunes' — an eerie booming drone emitted by some sand dunes — is stoking the controversy fuelled by rival theories.

Research on this striking natural phenomenon became something of a battleground after two groups, previously collaborators, put forward two opposing theories. Now, a team at the California Institute of Technology (Caltech) in Pasadena, led by mechanical engineer Melany Hunt, says that they are both wrong¹.

"There are strong feelings in this field," says physicist Michael Bretz at the University of Michigan in Ann Arbor. "It'll take a while longer to get it sorted out. But the explanations keep getting better."

The noise of the 'singing' dunes can be very loud, audible for up to ten kilometres. "It's really magnificent," says physicist Stéphane Douady at the Ecole Normale Supérieure in Paris, who proposed one of the competing theories. Marco Polo described it on his journeys through the Gobi desert in the thirteenth century, attributing the sound to evil desert spirits. There is a sand-covered hill in northwestern China called Mingsha Shan, which means singing sand mountain.

The effect is clearly related to avalanches of sand, and can be triggered by people sliding down the slopes. One of the first attempts at a scientific explanation came from Ralph Bagnold,

an army engineer who fell in love with the North African deserts during the Second World War. He suggested that the noise was caused by sand grains colliding, and that the frequency of the sound was determined by the average time between collisions. This implies that the frequency depends on the size of the individual grains, increasing as the grains get smaller.

Douady and his students Bruno Andreotti and Pascal Hersen focused on these collisions during a research trip in Morocco in 2001. Douady decided that for the moving grains to generate a single sound frequency, their motions must

become synchronized. This synchronization, he argued, comes from standing waves set up in the sliding layer. The noise is loud because the surface of the

dune acts like a giant loudspeaker membrane.

But Andreotti came up with a slightly different explanation. The synchronization of grain motions, he said, comes from waves in the sand below the sliding layer, which then act back on the moving grains, 'locking' their movements together.

It might seem like a small distinction, but Douady and Andreotti could not resolve their differences, and in the end they published separate papers offering their explanations^{2,3}. Andreotti now works at a separate lab in Paris.

But both explanations have serious problems, according to Hunt. For one thing, the

"You can take a cupful of this sand and excite it with your finger."



measurements made by her team on booming dunes in Nevada and California seem to show that the booming frequency doesn't depend on the grain size at all. What's more, the previous theories imply that all dunes should be able to 'sing' when an avalanche takes place. But in fact not all dunes sing — that's why Mingsha Shan got its name, for example.

Crashing DNA's ultraconservative party

A colony of mice whose very existence defies logic could rewrite our understanding of human evolution, health and disease, researchers say.

The laboratory mice lack stretches of DNA that scientists believed were essential for survival. And yet they eat, grow and reproduce normally. There seems to be nothing wrong with them despite their genetic deficiencies, says Nadav Ahituv, a human geneticist at the University of California, San Francisco, who created them through two painstaking years of breeding experiments. "We were expecting to see either fertility problems or a lethal phenotype"

— the mice being affected so severely that they die in the womb. But neither happened. "I was very surprised," says Ahituv, who published his results this week¹.

Ahituv made four mouse 'knock-outs', each one lacking a stretch of DNA between 222 and 731 base pairs long. These same stretches of DNA exist in human genomes, base pair for base pair. This 'ultraconserved' DNA is exactly the same across the long evolutionary distance between humans and mice and rats. So why the mice lived could answer fundamental questions about evolution.

Ultraconserved DNA was first described in May 2004, when a

group led by David Haussler at the University of California, Santa Cruz, reported the existence of 481 stretches of DNA more than 200 base pairs long with completely identical sequences in mice, rats and humans². Of these stretches, 256 were found in non-coding DNA, often far away from genes. Those found close to genes were often near developmental regulators, which hinted that the ultraconserved DNA might act as a switch in the cascade of events that shapes an embryo into a fully formed mouse or human. Beyond that, it wasn't clear what this class of DNA did, and so such regions have been called 'junk' DNA.

But the fact that it was preserved through evolution led people to believe it was important. Haussler's group calculated that the probability of just one of these elements popping up at random in the human genome was less than 10^{-22} . And on 17 August, Haussler's group reported in *Science* that natural selection processes are actively preventing ultraconserved DNA from changing. Haussler's group sequenced hundreds of pieces of ultraconserved DNA from 72 people and statistically analysed the pattern of variations in the ultraconserved regions. The pattern was consistent with that expected if natural selection was discriminating against mutations

**SMOKE SIGNALS**

Cigarette habit leaves genetic hallmarks.

www.nature.com/news



of the seismic waves increased from 180 metres per second to 310 metres per second.

The Caltech researchers think that this layered structure enables the surface to act as a kind of waveguide for acoustic energy, rather like the way an optical fibre channels light. So although they agree that the boom is transmitted to the air by a loudspeaker effect of the dune surface, they think that the frequency is set by the width of the waveguide layer of sand. Dunes that do not have this layered structure, such as smaller dunes, do not sing at all.

This is unlikely to be the last word on the matter, however. For one thing, the strange properties of the sand in 'booming dunes' are not just found in large structures. "You can take a cupful of this sand and excite it with your finger," says Peter Haff, a geologist at Duke University in Durham, North Carolina. "You can feel it vibrating, like running your finger over a washboard. But you can take sand from other parts of the dune, and there's nothing you can do to make it vibrate." Haff says that, although these theories may offer part of the answer, "there must be something else going on at a small scale".

Douady agrees. "The problem for the Caltech theory is that we can recreate these sounds in the lab," he says. He thinks that the sand layering might play a role in modifying the sound, but that it is "just a decoration" to the basic mechanism of booming.

Philip Ball

1. Vriend, N. M. *et al. Geophys. Res. Lett.* **34**, L16306 (2007).
2. Andreotti, B. *Phys. Rev. Lett.* **93**, 238001 (2004).
3. Douady, S. *et al. Phys. Rev. Lett.* **97**, 018002 (2006).

Sliding scale: the sound of dune avalanches could depend on their depth.

Andreotti proposed that 'silent' dunes aren't dry enough, or have grains of the wrong shape. But Hunt and her colleagues think that the answer lies deeper than this. They say that dunes have to be covered in distinct layers of sand to create a boom. Their measurements of

vibrations in the sand — made with an array of 'geophones' like those used to monitor seismic waves in earthquake studies — showed that the speed of these seismic waves increases in abrupt steps as the sand gets deeper. At 1.5 metres below the surface of one dune, for instance, the speed

in ultraconserved regions³. "This means there is a force preventing mutations in the ultraconserved regions from spreading throughout the general population," Haussler says. "What surprised me was the strength of the selection" — about three times as strong as selection on protein-coding regions.

However, although ultraconserved DNA seems crucial, scientists still aren't sure what it does or where it comes from. Ahituv was part of one group, headed up by Edward Rubin and based at Lawrence Berkeley National Laboratory, California, that found that 45% of 167 ultraconserved elements tested in developing mice acted as enhancers — regulators that tune up gene activity⁴. Scientists led by Steven Brenner at the

University of California, Berkeley, have found another regulatory role for ultraconserved elements in a family of human genes that regulate themselves by destroying their own messenger RNA templates⁵.

Only a single ultraconserved element has so far revealed its origins. By scanning genome data, Haussler's group found that one human ultraconserved element is 80% similar to a piece of DNA found in a 400-million-year-old class of ancient fish that includes the coelacanth⁶. The element had been shuttled into the fish genome by a genetic invader called a retroposon, but mammals have now co-opted it to boost expression of a brain-

development gene called *ISL1*.

Although little is known about the history of ultraconserved DNA, it should still be possible to determine what makes it so important that it has been kept unchanged for

"We were expecting to see either fertility problems or a lethal phenotype."

millions of years. "The ultraconserved region could play a role in human diseases and we are now deciphering this unexpected involvement," says

George Calin at the University of Texas MD Anderson Cancer Center in Houston. Ultraconserved DNA expression may be different in cancer cells and healthy cells, for example. Calin's findings are due to be published next week.

Ahituv's new results seem to contradict the little that is known

about ultraconserved DNA, says Haussler. He thinks that the knock-outs may have produced effects so small that they weren't obvious in controlled laboratory conditions. "Evolution has performed vastly more trials than can ever be performed in a laboratory," Haussler says. "It's possible that there's a pretty small effect that is difficult to measure in the lab, but is significant in the long run."

Erika Check

1. Ahituv, N. *et al. PLoS Biol.* **5**, e234 (2007).
2. Bejerano, G. *et al. Science* **304**, 1321-1325 (2004).
3. Katzman, S. *et al. Science* **317**, 915 (2007).
4. Pennacchio, L. A. *et al. Nature* **444**, 499-502 (2006).
5. Lareau, L. F. *et al. Nature* **446**, 926-929 (2007).
6. Bejerano, G. *et al. Nature* **441**, 87-90 (2006).

See Editorial, page 1.

Chinese law aims to quell fear of failure

BEIJING

In a bid to encourage scientific risk-taking and to curb rampant misconduct, China is planning to legislate for failure. A law proposed by the Ministry of Science and Technology (MOST) would allow Chinese scientists to report failures in their research without jeopardizing their chances of future funding. But critics say that policies with real impact, such as establishing systems for effective evaluation and oversight, are needed more urgently.

"Chinese researchers are under tremendous pressure to produce results," says Cheng Guodong, a geoscientist at the Cold and Arid Regions Environmental and Engineering Research Institute in Lanzhou, Gansu province. "Their performance in the lab has a domino effect on other aspects of their life, such as promotion, salary and social benefits. So the stakes are high if they fail."

Some believe that a consequent fear of failure is preventing Chinese scientists from taking risks, and results in research misconduct across the country. According to the draft law, failure of research projects would not have any professional bearing on researchers as long as they could demonstrate that the experiments were well thought out and properly conducted, says Xue Lan, executive vice-president of the Development Research Academy for the 21st Century at Tsinghua University, Beijing, who heads the team that has drafted the proposal.

It was announced on 26 August, the day before the international Organisation for Economic Co-operation and Development released a report calling for improved innovation capacity in China. In 2005, China became the world's sixth biggest spender on research and development, and it currently has more researchers than any country apart from the United States. But the report, commissioned by MOST, concludes that China has a long way to go in building the mature, national innovation system it wants.

Some commentators say that China needs to introduce radical reforms to improve its innovation capacity. But is the proposed legislation for failure the right way of going about it? Li Gong, chief executive of Mozilla Online, the Beijing subsidiary of the global software company, is not convinced. "The starting point of the proposal is fundamentally flawed. The definition of failure in research is problematic because negative results also yield knowledge," he points out.

"Risk-taking is the most important aspect in science."



LOU LINWEI/ALAMY

Under pressure: the push for results has left many Chinese scientists unwilling to do risky research.

"This is a useful but small step," says Muming Poo, a neuroscientist at the University of California, Berkeley, and director of the Institute of Neuroscience in Shanghai. But the proposed amendment is built on the premise that there is critical appraisal of a project's success or failure in China, which is not the case, he says. Many researchers believe that the lack of a transparent evaluation system is the real reason for the poor performance of China's scientists and that this needs to be tackled urgently. "Without that, everything will fall apart," says Cheng.

"Fear of failure is only partly responsible for rampant research misconduct in China,"

says Bai Lu, a neuroscientist at the National Institute of Child Health and Human Development in Bethesda, Maryland. "At the heart of the matter is that there is no effective oversight mechanism to ensure that fraudsters do not go unpunished. The lack of a culture of accountability is a much broader and deeper issue."

According to the newspaper *China Daily*, the draft law says that principal investigators should establish credibility records for their staff in order to curb fraud, and asks researchers to be 'self-disciplined'. "But self-discipline is not enough," says Lu. "That's the point."

Despite the criticisms, the proposed law has its supporters. Duan Yibing, a science-policy researcher at the Institute of Policy and Management in Beijing, explains that basic

laws carry supreme authority in China, so legislation that recognizes the intrinsic risks associated with research and is tolerant of failure would send a powerful signal.

"Risk-taking is the most important aspect in science, so China is doing the right thing," says Raghunath Mashelkar, president of the Indian National Science Academy in New Delhi. "The law would influence the way researchers and funding agencies behave and generate an environment conducive to innovation." In India, Mashelkar has successfully carved out a niche for innovation by setting up special funds, such as the Kite-Flying Fund and the New Millennium Indian Technology Leadership Initiative, for high-risk projects, a strategy he thinks may offer China some useful lessons.

Legislation alone would not be enough to cause fundamental change in the culture of scientific research in China, acknowledges an employee at MOST, who wishes to remain anonymous. He says that a series of supporting regulations and policies, including those for improving appraisal and oversight systems, are already under discussion. "The law would not only provide a guiding principle for formulating those rules but a significant impetus for their implementation," he says.

The proposed law is being deliberated by the Standing Committee of the National People's Congress, China's top legislature. A decision is expected to be made next March.

Jane Qiu

Axe-wielding professor fined for dumping ether

Daniel Storm, a pharmacology professor at the University of Washington in Seattle, just wanted to save a buck. After being told that it would cost his lab \$15,000 to dispose of five containers of ethyl ether properly, Storm broke the three made from metal open with an axe in June 2006 and clandestinely poured the contents of all five down the drain.

But faked disposal paperwork tipped off safety inspectors. On 28 August, in a federal court in Seattle, Storm was sentenced to 80 hours of community service and given a \$5,000 fine. Prosecutors said that his manoeuvres with the axe could have caused a spark and ignited the ethyl ether, which is extremely flammable both as a liquid and as a vapour.

Hybrid embryos win support from UK public

Most of the British public back the creation of hybrid embryos consisting of human DNA implanted into animal eggs, if it will benefit medical research, a public survey suggests. The survey of more than 2,000 people, carried out for the Human Fertilisation and Embryology Authority (HFEA), found that 61% were in favour of the procedure if it might help scientists to understand diseases such as Parkinson's. A quarter of the respondents were against the technology.

As *Nature* went to press, the HFEA was widely expected to announce its approval of the technique at a meeting on 5 September to discuss the results of the public consultation. That should pave

the way for the granting of licences to two British research groups that applied in November last year for permission to create such embryos. For the latest details on the decision, visit www.nature.com/news.

Green light for centralized bank of patients' gene data

The US National Institutes of Health (NIH) announced plans on 22 August to store patients' genetic and medical information in a central bank. The policy will mean that NIH-funded researchers must submit genome-wide scans for diseases into a database from 2008. Other researchers can access the data by request.

The NIH revised the policy from an earlier draft issued in August 2006, but made few changes despite public comments citing concerns over privacy. The database won't include information to identify patients, says Elizabeth Nabel, director of the National Heart, Lung, and Blood Institute, who chaired the committee that created the policy.

Scientists who gather the gene data have exclusive publication rights for 12 months. The policy does not preclude commercialization of links between genes and diseases, but it urges researchers and companies to patent only new applications, such as diagnostic tests.

Space-lab staff file lawsuit to protect their privacy

Twenty-eight scientists at the Jet Propulsion Laboratory (JPL) near Pasadena, California, have filed a federal class-action lawsuit



NASA/JPL-CALTECH

Staff at the Jet Propulsion Laboratory are up in arms over new rules on background checks.

against background checks that they see as violations of privacy.

All 5,000 employees at JPL face a 27 October deadline for complying with a Bush administration directive, which requires fingerprinting and extensive background checks before employees can receive new identification badges. The researchers who are suing — including several of JPL's most senior scientists — object to open-ended probes of personal issues, including sexual preference.

Attorneys for the scientists are seeking a preliminary injunction to halt the checks. The suit is against NASA, the US Department of Commerce, and the California Institute of Technology in Pasadena, which employs and manages JPL staff.

Publishers campaign against open access

The Association of American Publishers* is taking part in an initiative to protest against what it calls government interference in the scholarly communication process.

Some groups and legislators are pushing for all publicly financed research to be made freely available to the public. Many traditional publishers object, and some have used aggressive tactics to fight the movement (see *Nature* 445, 347; 2007).

The initiative — called the Partnership for Research Integrity in Science and Medicine — says that it wants to provide the public with more information about scholarly publishing. One of its principles is that "society is best served by sustainable business models and reasonable copyright protections". News of the group's formation did not go down well in the blogosphere, where a number of critics attacked it for implying that open-access publication harms peer review.

**Nature's* US division, Nature America, is a member of the Association of American Publishers.

Biodiversity agency to sue over endangered species

Tired of one-off lawsuits over species it feels should be protected under the Endangered Species Act, a US conservation group has gone all out. The Center for Biological Diversity, based in Tucson, Arizona, has announced its intention to sue the US government in an enormous lawsuit covering 55 plant and animal species, including Florida manatees (pictured).

The move comes in response to claims that a now-ousted government official, Julie MacDonald of the Department of the Interior, acted as a lone wolf in meddling with agency scientific reports about which species should be listed or de-listed for protection. "The idea that she was this singular bad apple that has been excised is ridiculous," says Kieran Suckling, policy director for the conservation group.

The suit will go after an alleged systematic pattern of scientists being illegally overruled by political officials. "I cannot comment on pending legal matters, but I can tell you that the Department of the Interior takes very seriously its responsibilities under the Endangered Species Act," says Chris Paolino, a spokesman for the department.



D. PERRINE/NATUREPL.COM

BUSINESS

King coal constrained

Sustained high oil prices won't be enough to make coal liquefaction economically viable without large-scale public investment.

Katharine Sanderson reports.

Turning dirty coal into a clean-burning liquid fuel remains something of a challenge for the energy industry. As scientists heard last month at the annual meeting of the American Chemical Society in Boston, Massachusetts, the process behind coal liquefaction might be well known — but its large-scale adaptation could still be some time off.

"The economics at this point aren't very useful," says Caroline Burgess Clifford, a fuel chemist at Pennsylvania State University's Energy Institute in University Park. "We don't have a lot of data for people to rely on." And greenhouse-gas emissions are also putting investors off. "People are less likely to invest in coal liquefaction because of the carbon dioxide problem," she says.

Coal's long-standing association with air pollution and CO₂ emissions has deterred many observers from taking it seriously as a 'green' fuel. Coal liquefaction has traditionally been a route taken by countries — such as Germany in the 1930s and South Africa later on — that simply have no other option.

Yet with oil prices showing no signs of dropping, governments and energy companies are showing a renewed interest in coal — encouraging scientists and engineers to revisit the technologies for retrieving petroleum-like liquid fuels.

The predominant coal liquefaction process, developed by Franz Fischer and Hans Tropsch at the Kaiser Wilhelm Institute for Coal Research in 1923, uses high temperatures and pressures to break down the long hydrocarbon chains in coal, forming a mixture of carbon monoxide and hydrogen. This is passed over an iron- or cobalt-based catalyst to produce a liquid that contains hydrocarbons of various lengths. The liquid can then be refined in the same way as crude oil (see *Nature* **444**, 677–678; 2006).

The technique was developed further by the energy company Sasol in Johannesburg, South



Economic isolation forced South Africa to refine the coal technology used today at Secunda, Mpumalanga.

Africa. Sasol was attempting to meet the country's energy needs when South Africa, with its apartheid regime, faced an economic embargo that halted its oil imports. The United States and other nations also explored the technology during the oil crisis in the mid-1970s but soon after that, global research and development into coal liquefaction came to a virtual standstill.

Direct approach

The Fischer–Tropsch process is indirect, producing a gas mixture that is then recombined into liquid hydrocarbons. Direct approaches have also been proven, but on a smaller scale. These involve pulverizing coal and mixing it in a solvent, which is then treated with hydrogen at 450 °C, at high pressure and in the presence of a catalyst. As in indirect methods, the mixture can then be refined in a similar way to oil.

Elliot Kennel, a chemical engineer at West Virginia University in Morgantown, says that direct liquefaction is less economical than the Fischer–Tropsch process because it consumes large amounts of hydrogen gas, which is expensive to produce.

So Kennel is working to reduce the amount of hydrogen needed, from the 10%, by mass, needed for direct methods to produce a light petroleum, to as little as 0.5%. This will also cut the energy needed to drive the process. "To hydrogenate coal, to absorb hydrogen, takes energy," he explains, "If we can absorb less hydrogen we can expend less energy."

Kennel's process produces a thicker, heavier mixture of hydrocarbons than other direct methods, which will yield less petrol in the refinery and more thick, tar-like residues. But

he says that the by-products are low in sulphur and heavy metals, making them easy to sell. Kennel's team is partnered with several industrial consumers of these heavy products, including GrafTech, a leading graphite manufacturer based in Parma, Ohio, and Alcoa, an aluminium producer based in Pittsburgh, Pennsylvania.

Burgess Clifford is also working towards a more attractive process for liquefaction. She has, for example, run tests at a small pilot plant to produce a hydrocarbon mixture rich in the heavy oils used to fuel aircraft. The process blends coal tar with light oil — a distillation product from the refinery — and hydrogenates the mixture using catalysts and conditions typical of those used in a refinery. But she says that the process is not yet ready for commercialization. "We could implement this immediately," says Burgess Clifford, "but it wouldn't be what a refinery would want."

Tentative steps

Industrial companies have joined forces to build a \$600-million pilot plant in Gilberton, Pennsylvania. The plant will use the Fischer–Tropsch process to convert about 4,300 tonnes of coal a day into transportation fuels, heat, electricity and various chemical by-products. The project is being supported by the US National Energy Technology Laboratory and involves Sasol, Shell, Californian energy consultancy Nexant, and Uhde, a chemical-engineering company based in Dortmund, Germany. The companies say that they will build more plants if the pilot is successful.

And in China, several commercial-scale coal-liquefaction projects are under way, including one direct liquefaction project started in 2004 by the Shanghai-based coal company Shenghua in Erdos, Inner Mongolia,

"People are less likely to invest in coal liquefaction because of the carbon dioxide problem."

— Caroline Burgess Clifford



SASOL

which, when it opens in 2008, will convert 6,400 tonnes a day of coal into oil products in its first year, with plans to increase that capacity over the following years.

Kennel says that the principles of coal liquefaction are well established, and that the emphasis is on refining the existing processes. "Today we're talking about making them cheaper, more efficient and more environmentally friendly," he says.

Capturing the market

This last point is really the elephant in the room for coal liquefaction. Because processing a tonne of coal produces much more CO₂ than a tonne of crude oil, the environmental case rests on the hazy prospect that the CO₂ can be 'captured' or sequestered. None of the pilots under construction will achieve that.

"The thing on everybody's mind is CO₂," admits Kennel. "The community is trying to figure out how to capture it." Clifford says that concerns about CO₂ emissions are preventing major commercial investment in coal liquefaction, and that public money is needed if new plants are to be built.

Sasol claims that the CO₂ from its plants is purer than that produced by coal-fired power stations, and it should therefore be easier to liquefy and sequester. But that will be easier said than done. And most people outside the coal industry are sceptical about the economic feasibility of coal liquefaction — even before the large and unknown costs of sequestration and storage are factored in.

High oil prices certainly make technologies such as coal liquefaction more viable, says Robert Wine, a spokesman for the oil company BP in London. "They are expensive projects," he says. "They can only compete either when oil prices are high, or when they can attract government support."

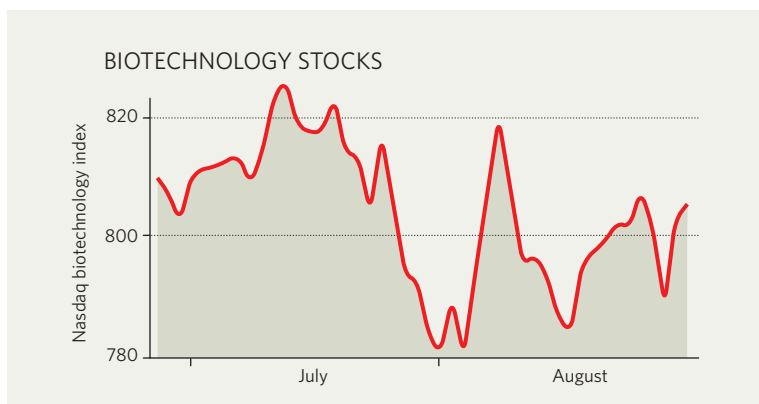
IN BRIEF

SECURE VENTURES Companies with interests in homeland security are becoming the hottest ticket for support from venture capital. London-based Venture Business Research says that security-related companies have attracted US\$4.5 billion in venture capital and private equity investment so far this year, up 34% on the same period in 2006. It adds that the sector is now vying with the clean-energy sector, which has attracted \$5.2 billion this year, as the most popular destination for venture capital worldwide.

ENERGIZED The British government has selected defence contractor QinetiQ in Hampshire, UK, to run a body that will coordinate research into energy flow in materials. The Ministry of Defence contract to the UK Energetics Consortium, as the new outfit is called, is expected to be worth £10 million (US\$20 million) over three years. The ministry says that it needs the consortium to help it garner knowledge about energetics from universities, companies and government laboratories.

BOEING BOOST NASA has awarded a US\$515-million contract to Boeing to build part of the rocket that will be the space agency's main astronaut launcher once the space shuttle is withdrawn from service in 2010. Boeing will build the upper stage of the Ares I rocket, which will carry astronauts to the International Space Station and, perhaps, to the Moon. Alliant Techsystems of Edina, Minnesota, has already won a larger contract to build the rocket's lower stage.

MARKET WATCH



This week, Wood Mackenzie, an Edinburgh-based research and consulting firm, reviews recent trends in biotechnology stocks.

The Nasdaq Biotechnology Index made no progress during July and August, and stands only marginally higher than it did at the start of the year. Broader market indices lost value during the period.

Bad news continued for California-based Amgen over this period. On 30 July, US reimbursement authorities issued a ruling that will reduce the use of erythropoietin for treating anaemia in patients with cancer in the Medicare and Medicaid programmes. That adds to the pressures building up on Amgen's flagship erythropoietin drug, Aranesp, which is embroiled in safety concerns and losing sales (see *Nature* **448**, 121; 2007).

Then on 15 August, Amgen announced a 14% reduction in its workforce — much of which is likely to be in its research-and-development staff. The

company's share price has continued to fall. Over the past two months it has dropped nearly 10%; its value is now at its lowest point for more than four years.

Amgen's misfortunes were offset by strong performances by other companies in the index. At the end of July, when reporting its second quarter results, Vertex of Cambridge, Massachusetts, announced good results from clinical trials of its experimental hepatitis C drug, telaprevir — widely viewed as a potential blockbuster. The company's share value climbed by around 35% over the period.

The stock of Onyx Pharmaceuticals in Emeryville, California, has also climbed steadily. In August, its marketing partner Bayer reported stronger-than-expected sales of its newly approved kidney-cancer drug Nexavar (sorafenib). The drug also achieved impressive clinical results for the treatment of liver cancer, potentially broadening its market.

YEAR OF THE TIGER

Dubious science and looming legalization of the tiger trade threaten to derail China's efforts to save the Siberian tiger. **Jerry Guo** goes to the world's largest tiger-breeding facility to investigate.



The automated gates chug and clatter open as a jeep, its windows ribbed with steel, noisily announces its arrival in the tiger park. Without the usual gaggle of tourists to impress, the occupants of a neighbouring jeep toss out a skinny pheasant as the driver shouts obscenities at a dozen lounging Siberian tigers. For sightseers they would have released a bull, but they cost US\$250 each.

One tiger finally takes notice and lunges at the fluttering fowl, which has enough brains to scuttle under one of the jeeps. The tiger, neither as sharp nor as small as the pheasant, slams into the vehicle with a thud. And as the hulking beast shakes off the dust and disappointment of his failed attempt, the pheasant dashes into the brush. The striped leviathan promptly settles back down, seemingly deciding that the prey isn't worth the effort.

And why not, for these tigers are already well-fed, particularly by the 300,000 tourists who flock every year to the tiger park at the Hengdaohezi Feline Breeding Centre on the outskirts of Harbin in northeastern China's Heilongjiang province. By most accounts, the place is an enviable success. Started in 1986 with 8 Siberian tigers, it is now home to 800 of the big cats. Compare that with the estimated 150 Siberian tigers in US zoos. The largest tiger-breeding facility in the world, Hengdaohezi — like its cousin down south at the Wolong Panda Reserve — has learned the art of churning out cubs, 100 this year alone.

But this year the centre has been subject to all

sorts of media attention, from gruesome videos on the Internet of tigers eviscerating a bull as tourists gape, to reports of plans to reintroduce 600 of the cattle-fed, people-friendly tigers into the wild. At the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) meeting this June in The Hague, the Netherlands, China's tiger-breeding programme was criticized for at best creating tourist traps, and at worst being flat-out farms for the animals. Indeed the Hengdaohezi facility was developed as a government-run enterprise to capitalize on the tiger-bone trade. Since the practice was outlawed in 1993, the park has depended on tourism for 80% of its \$4 million per year operating costs. But shortly after the CITES meeting, Wang Wei, a deputy director at the State Forestry Administration in Beijing threatened the imminent reopening of the tiger trade, inviting 70 international tiger experts to Hengdaohezi in July to hear the merits of such a move (see 'Another pickle for Siberian tigers').

So far, Western scientists are unconvinced that the proceeds from farming the animals might fund conservation efforts. Moreover they doubt whether conservation is something the facility is interested in or even equipped to do. "They want to use their bones and parts," argues Sybille Klenzendorf, who toured the facility in 2006 as director of species conservation for

the conservation group WWF. "It's basically a tiger farm operating under the pretence of a research facility."

In a concrete bunker off-limits to gawking tourists, mother tigers nurse their cubs in tiny cages. The park's chief scientist, Liu Dan, proudly surveys his charges. For him, the objective is straightforward. "Our goal is to reintroduce them into the wild," he says. He denies media reports of an earlier failed reintroduction. The centre did, however,

send ten tigers to a small area resembling alpine forest in the Changbaishan reserve, close to the North Korean border. "It's a very good wild habitat. A good exercise in all aspects of training, but still a big difference to the wild," says Liu Dan.

Genetic sleuths

The park contains roughly twice the number of Siberian tigers that exist in the wild, and letting loose even a few captives would have widespread conservation implications — especially in the small remaining natural range in northeastern China where perhaps ten tigers reside. But reintroduction wouldn't be just about bolstering the wild population.

In 2004, Michael Russello and his colleagues at Yale University in New Haven, Connecticut, published a study in *Conservation Genetics* indicating that all but 2 of the roughly 60 wild Siberian tigers they sampled shared a single

"It's basically a tiger farm operating under the pretence of a research facility."

— Sybille Klenzendorf

E. DALZIEL/AP

CHINA DAILY/REUTERS

mitochondrial haplotype¹ — a set of genes that is inherited en masse. “The genetic diversity was about as low as it gets,” Russello says. In particular, his data suggested that captive tigers, at least those in North America, may be more genetically diverse than their wild counterparts. But he doesn’t know if this corollary holds true for Hengdaohezi’s 800 tigers. If so, “they could re-inject variation that has been lost in the wild”.

At Wolong Panda Reserve, keepers are increasing the population to maintain a healthy genetic reservoir in case of a sharp drop or extinction in the wild. Three hundred pandas is apparently the magic number, and tourists are no less impressed. There, as in Hengdaohezi, even keeping the animals caged can benefit conservation, as long as pedigrees are tracked and specific pairs matched to maximize diversity, says Shujin Luo, a conservation biologist at the National Cancer Institute in Frederick, Maryland. That’s not the case at Hengdaohezi. Although Luo is currently analysing the genetic diversity of wild versus captive Siberian tigers, like Russello and other Western researchers, she has not been able to obtain any samples from Hengdaohezi.

One person who has obtained samples says he is confident in the diversity of the Hengdaohezi stock. Across town from the facility, forensic geneticist Xu Yanchun of Northeast Forestry University in Harbin, has collected 500 blood samples from the place and plans to have a completed ‘genebank’ of the captive population by this winter. “I’m sure genetic diversity here is higher than in the wild,” he argues, citing indicators such as heterozygosity and allelic distribution from his unpublished data. “This population is quite high in genetic diversity because they are well managed,” Xu says.

Fuzzy breeding

The tourists love the liger enclosure — they can’t snap enough pictures as the tour bus slowly rolls past lions, tigers and their enormous hybrid offspring, all basking next to each other. The huge animals, a cross between a male lion and a female tiger, are a dramatic sight, but such disregard for intermixing could lead to bigger problems. The property also contains Bengal tigers, technically a different subspecies from their Siberian cohabitants, and the subspecies could produce harder-to-spot hybrids together.

In general, Hengdaohezi’s breeding strategy is crude compared with Western practices. Unlike US and European zoos that use computer models to calculate exactly which animals should mate together — and stud books to track every individual at Hengdaohezi, Xu’s genetic pedigrees are mostly ignored. Liu Dan concedes that the centre doesn’t control or record which tigers breed together (as long as they’re not brother

and sister). “We don’t have the resources,” he says.

“Now they’re all breeding haphazardly,” says Liu Yutang, a cryogeneticist at Northeast Forestry University. “We have to wait until our technology is better so that we can control which tigers will mate.” In captivity each female may mate with several male tigers when sexually receptive, confusing keepers on the paternity of the resulting cubs. The centre also rarely exchanges tigers with other breeding facilities. US zoos regularly shuttle tigers across the country to breed, explains Kathy Traylor-Holzer, the Siberian tiger stud-book keeper for American zoos. “The fear is to accept an animal from Harbin as they may carry genes from other subspecies,” she says. “If you don’t manage the population, you automatically lose genetic diversity.”

The breeding facility is not a member of the international stud book for Siberian tigers, which includes almost all US and European zoos. “It’s because they breed the same animals over and over again, which you would never do as a registered stud-book zoo,” says Klenzendorf. Xu, who serves as a breeding consultant to Hengdaohezi, refuses to use the international stud book’s standard software program, SPARKS, which calculates the degree of inbreeding for each individual. “The prediction is not accurate,” he says, citing more unpublished data that indicate that sperm genetically similar to the female’s genotype stand less of a chance in the oviduct, a case of ‘selective fertilization’. “This model has the precondition that all alleles are passed on randomly, which is not accurate because of my selectivity theory,” says Xu.

But Xu’s claims about the population’s high genetic diversity draw doubts even from



Killer moves: tigers in the Hengdaohezi park learn to hunt by killing livestock, in what is often a display for tourists.



colleagues such as Liu Yutang who says physiological problems from inbreeding already run rampant. “Some tigers are very weak, have different stripes, high cub mortality rates, and reduced immune systems,” he says. “And when they’re all related, then it’s easier for something to wipe them all out,” he adds, citing the waves of deaths from bird flu and canine parvovirus. Liu Dan says he’s not aware of these tiger deaths, but Xu says they numbered in the “several tens” in 2005.

Luo says she reviewed one of Xu’s papers on the genetic fitness of Hengdaohezi tigers and didn’t consider it good enough for publication. “Every few months we receive papers from China about tigers; overall the quality is disappointing,” she says. So far, Xu’s data from the past decade at Hengdaohezi have not been published in a Western journal, although he did publish a paper in *Forensic Science International* detailing a genetic fingerprinting method to combat Siberian tiger poaching².

Siberian sperm bank

In his university office, Liu Yutang grabs a giant syringe equipped with a video camera and a light. To help bring the centre to Western standards by controlling exactly which tigers mate, he has developed this gadget for artificially inseminating the females. He tried it earlier this year with no luck. Still it was an

T. O’ROURKE/SINOPIX/REX FEATURES

Another pickle for Siberian tigers

On a nondescript street near downtown Harbin, the Double Mountain Local Products Wholesale Centre offers the usual array of kitsch items stripped from the wilderness: deer antlers, pelts and dried starfish. A request for tiger wine, a traditional brew of corpse-steeped cheap liquor with dozens of reputed medical benefits, raises a stern eyebrow from an employee who informs a customer that as such concoctions are illegal, they are not available at the store. But at the mention of American money, a store manager intervenes — \$100 would buy two bottles, and true to the employee's words they are not at the store; they will be delivered via courier. Doubts about the brew's authenticity are shooed away. The manager is certain the bottles are the genuine article because, she says, "they came from over at that tiger park". She is referring to the Hengdaohezi Feline Breeding Centre on the outskirts of the city. And whether or not she is speaking the truth, the manager is highlighting a looming international stand-off between

conservationists and the Chinese government.

China banned domestic trade of tiger parts in 1993, but that did not arrest the desire for their use in wine or traditional Chinese medicine. A black market fills the demand and goods can often be traced back to breeding centres. In August 2006, a tiger farm in Guangxi province was caught with 400 vats of wine, each stewing a whole tiger carcass. This June at the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) conference in The Hague, the Netherlands, wildlife officials used DNA evidence to accuse the same farm of serving tiger meat.

In a walk-in fridge at Hengdaohezi — off-limits to tourists and journalists — 200 frozen tiger carcasses lie scattered, waiting to be turned into tiger wine and medicine, according to Xu Yanchun, a breeding consultant for the park at neighbouring Northeast Forestry University. Whether Hengdaohezi benefits tiger conservation is questionable, but one thing is



Dark liquor: is it tiger wine?

certain — if the government lifts the ban on the tiger trade, places such as Hengdaohezi will profit.

Liu Dan, the park's chief scientist, doesn't see a problem. "We can use dead tigers to save live tigers," he explains, promising to use profits for the centre's genetic and reintroduction projects. And the government seems to agree. "It's very hard to go against these pressures to open the trade," says Wang Weisheng, manager

of wildlife resources at the State Forestry Administration.

Wang says a decision on the ban could be made as early as October. As of 2006, all tigers have been required to wear a microchip, and Wang says such tracking abilities combined with a certification process — a system that met with success with China's ivory, crocodile and ginseng trade — could lead to a win-win situation for everybody. But lifting the ban may be illegal. Craig Hoover, chief US CITES enforcement officer, says China would be flaunting an existing international ban on tiger parts — and noncompliance could lead to sanctions.

Although several Western economists suggest harvesting captive tigers would relieve poaching pressures, and ultimately funnel money to conservation efforts, most conservationists disagree. "You can't possibly saturate the market with just parts from tiger farms," says John Goodrich, a conservationist with the Wildlife Conservation Society, New York. "It'll be devastating for wild tigers."

J.G.

improvement on the prototype, which, he says with a grimace, was "too sharp".

Liu Yutang says he is attempting to assemble a Siberian tiger sperm bank with samples from the entire Hengdaohezi population for artificial insemination. He's even eyeing wild tigers as possible targets for collection, but there's still a lot to learn. He admits his team doesn't yet know all of the basics, such as when tigers produce sperm. "A lot of this work is through trial and error," he says. For example, 8 of 23 semen samples he has extracted so far from the captive tigers yielded no sperm.

Liu Dan's plans to reintroduce tigers into the wild have faced further criticism. "They'll wreak havoc in the villages after being fed chickens and getting used to jumping on cars," says John Goodrich, a conservationist with the Wildlife Conservation Society in New York. "And there's absolutely no need to release tigers at all when tigers from Russia will move into China."

For now, the Siberian tiger's foothold seems sturdier than that of its cousin, the South China tiger. As a result of poor

breeding and poaching, the South China population now numbers 66, all caged in a handful of zoos. Xu says the existing population is extremely inbred, with high mortality and low fertility. A paper under review at *Current Biology* paints an even bleaker picture. Yue Bisong of Sichuan University in Chengdu, notes that of 45 tigers he sampled, only 13 were pure South China tigers, the rest were hybrids with other tiger subspecies.

With its baffling breeding techniques and



King of the jungle: Liu Dan oversees 800 Siberian tigers.

plans to open a market in tiger parts, Hengdaohezi hardly seems the safest place for Siberian tigers, but how they would fare in the wild is even more uncertain. So perhaps it is fortunate that the reintroduction campaign is mainly hype for now. Although media reports mention plans to release 600 of the captive tigers (apparently hoping to coincide it with the Beijing Olympics), the centre has not yet separated any group for eventual reintroduction, selected any potential release sites, or built specialist training enclosures. As Liu Dan broods over his nursing mothers, he defends the conservation work of the centre, posing the rhetorical question that if they weren't keeping the tigers around for a greater purpose, wouldn't they be just another tiger farm? "From breeding to reintroduction is a long process," Liu Dan says. "The programme isn't mature yet."

Jerry Guo is a freelance writer in New Haven, Connecticut, reporting from China.

1. Russello, M., Gladyshev, E., Miquelle, D. & Caccione, A. *Conserv. Gen.* **5**, 707-713 (2004).
2. Xu, Y. C. et al. *Foren. Sci. Int.* **151**, 45-51 (2005).

See Editorial, page 2.

TIME TRAPS

The whole world felt the effects of the dinosaur-killing mass extinction 65 million years ago. But a spot in Colorado may have the best record of it. **Rex Dalton** reports from Denver.



J. VRIESEN/DENVER MUSEUM OF NATURE & SCIENCE

It's no wonder palaeontology is so popular in Colorado: fossils practically permeate the landscape. Hiking trails run through hills known for *Triceratops*. Part of the local interstate is named after *Tyrannosaurus rex*. And around the east of the city of Denver you can drive out into the grassy plains and walk the landscape where dinosaurs once lumbered.

But for some, the thrill isn't just in finding the giants — it's in dating precisely when they disappeared. Here, in the rolling expanses east of the city, lies what could be the world's best spot for understanding what transpired 65 million years ago, when the Cretaceous geological period ended and the Tertiary one began. This 'K/T boundary' marks a crucial moment when some cataclysmic event — probably an asteroid impact (see page 48), or massive volcanic eruptions — wiped out most creatures worldwide.

The going date for this event is 65.5 million years ago, with a margin of error of 100,000 years¹. But one group of scientists is aiming to get this margin down to just 25,000 years. Their work is part of a major international geochronology effort called EARTHTIME, which aims to calibrate the geological timescale for

all of Earth's history². "We hope to use the hundreds of ash layers to develop an unprecedented time sequence that will allow us to read the geological record of deep time like never before," says palaeobotanist Kirk Johnson, leader of the team and head curator at the Denver Museum of Nature and Science.

Colorado has taken centre stage in geology after some serendipitous discoveries in the Denver basin over the past 15 years. In 1994, plant fossils of an ancient rainforest were unearthed as a highway was being built near the town of Castle Rock, about 40 kilometres

Artist's impression of West Bijou valley soon after the asteroid strike 65 million years ago.

south of Denver. Dating techniques pegged them at 64.1 million years old. That suggests that a tropical rainforest had established itself less than a million-and-a-half years after the cataclysmic extinction — earlier than some had suspected³.

More discoveries followed, defining the Cretaceous underbelly of Denver. Then in 2003, Johnson and his co-workers published a report⁴ that detailed the wealth of K/T sediments at a site called West Bijou Creek, about 65 kilometres east of Denver. Suddenly, Colorado took on world-class importance. Sediments marking the K/T boundary can be found across the globe, but the West Bijou site is one of the most complete locations, with plant and animal fossils encased within layers of volcanic ash.

In an eroded knob along a gully overlooking the grassy plains, Johnson points to a layer just below the surface that is rich in the element iridium, which is often used as an indicator of an extraterrestrial impact. Below the iridium layer marking the K/T boundary are dinosaur fossils; above it there are none.



The region's rich geological history makes it ideal for studying the events leading up to, at and just after the K/T extinction. Volcanoes once dotted the eastern side of the Front Range of the Rocky Mountains, showering ash over the countryside. Eventually, the volcanoes eroded away, but they left behind thick layers of sediment that piled up in the basin. It is these alternating white and dark layers that provide geologists with clues about the past.

On past field trips, Johnson and his crew scoured the sites for plant fossils — finding about 500 leaves and 2,000 pollen samples, which together show what the environment was like around the cataclysmic transition. The sizes of leaves can indicate rainfall⁵, and the edges of leaves denote temperature — the smoother the edge, the warmer the temperature must have been⁶.

"Leaves are a fossil thermometer and rain gauge," says Johnson. "They also are a chalk board for insect tracks." By studying insect markings on the leaves, scientists can see which species went extinct and when. For instance, the team found little damage due to insect feeding on leaves dated at 63.8 million years old — suggesting that food webs were still out of kilter between a million and 2 million years after the extinction event⁷.

Dating service

Now, though, the researchers are moving beyond fossils and into geochemical, magnetic and other evidence to help them improve the chronology of what happened across the K/T boundary. Johnson has been joined by Sam Bowring, a specialist in geochemical dating at the Massachusetts Institute of Technology in Cambridge and William Clyde, a geologist from the University of New Hampshire in Durham who analyses palaeomagnetic fields.



William Clyde checks the magnetic polarity of ash layers in the Denver Basin.



Kirk Johnson (left) and Sam Bowring trawl through sediment in search of zircon crystals.

Their quarry is tiny zircon crystals in the layers of ash. By measuring the ratio of uranium to lead in individual zircon crystals, Bowring can date the ash layers and help the team pinpoint the time at which the K/T extinction occurred.

Hiking to the bottom of a gorge called My Lucky Number, Johnson leads Bowring and Clyde to an eroded outcrop. Here they dig a shallow trench, revealing a dozen layers of ash. "You are looking at the surface of the swamp where the ash fell" about 500,000 years after the K/T boundary, says Johnson, sweeping clean the first whitish layer.

Every time a new layer of ash is uncovered, Bowring kneels — trowel in hand — and scoops out samples for the collection bag. "Oh, baby," is his frequent refrain. A field crew can work for years before finding a single ash layer. Here, more and more appear beneath the trowel.

Meanwhile, Clyde uses clues from the periodic reversals in Earth's magnetic field. He takes a wood rasp, flattens a section of the soft ash layer, and aligns a compass along its plane. Then he digs out a fist-sized piece and tucks it away in the collection bag. These pieces will later be tested to see whether they fall in the same polarity as occurs today, or in the 'reversed' direction. By working out the date of the reversals around the K/T layer, the team can narrow down the time at which the extinction event must have occurred.

Back in time

For the rest of the day the researchers work their way up the gorge, scaring off deer as they go. They repeat the exercise half-a-dozen times, eventually taking readings on some 40 layers of ash. As they sample the rising sediments, Clyde may find five consecutive normal palaeomagnetic readings, then five consecutive reversed ones. Bowring's zircon dates can perhaps be used to bracket and date that reversal point.

"Leaves are a fossil thermometer, rain gauge and chalk board for insect tracks."

— Kirk Johnson

By comparing the findings to a research core stored at the US Geological Survey's office in Denver, near-surface dates can be correlated with the same sediments found deeper at other locations. And by comparing the new Denver dates to records from the ocean cores, the researchers might even be able to improve the correlation of sediment layers found around

the globe. For instance, European researchers have found evidence⁸ in Spain for changes in Earth's orbital cycles near the K/T boundary, which they date to 65.8 million years ago. The Denver studies will test that result, and help to refine the chronology of the past climate

changes caused by changes in the Earth's orbit.

These observations will be plugged into the larger EARTHTIME project, which is just beginning to get under way. But already, Johnson is looking at short-term benefits of the research — in particular, tying it back to issues concerning the health and development of the Denver area.

His team's work on local aquifers, which run through sediments from the Cretaceous period, has helped to clarify the region's geology and hydrology. And that, in turn, could help city planners make better decisions about how to manage water supplies for the rapidly growing metropolis — to keep itself from going extinct. ■

Rex Dalton writes for *Nature* from San Diego.

1. Gradstein, F. M. et al. *A Geologic Time Scale 2004* (Cambridge Univ. Press, 2005).
2. Dalton, R. *Nature* **444**, 134–135 (2006).
3. Johnson, K. R. & Ellis, B. *Science* **296**, 2379–2383 (2002).
4. Barclay, R. S., Johnson, K. R., Betterton, W. J. & Dilcher, D. L. *Rocky Mt. Geol.* **38**, 45–71 (2003).
5. Wilf, P., Wing, S. L., Greenwood, D. R. & Greenwood, C. L. *Geology* **26**, 203–206 (1998).
6. Traiser, C., Klotz, S., Uhl, D. & Mosbrugger, V. *New Phytol.* **166**, 465–484 (2005).
7. Wilf, P., Labandeira, C. C., Johnson, K. R. & Ellis, B. *Science* **313**, 1112–1115 (2006).
8. Dinarès-Turell, J. et al. *Earth Planet. Sci. Lett.* **216**, 483–500 (2003).

See Editorial, page 1.

The collector

How **Paddy Patterson**, one of the architects of the Encyclopedia of Life, hopes to present biodiversity to the world.

David 'Paddy' Patterson started his career off small, working on the taxonomy of protists and protozoa and putting descriptions of them up on the web in projects such as Micro*scope, Microbial Life and the International Census of Marine Microbes. Now, the biologist, who works in Woods Hole, Massachusetts, is going macro in a big way with the Encyclopedia of Life (EOL), a project that intends to create a website for every known species on Earth.

So what is the proposed schedule for the EOL?

In February next year, hopefully, there will be a major release of the first edition of the EOL. The expectation is that within a ten-year period we will have relatively well-informed pages on all 1.8 million species.

How can you realistically hope to get that much done?

The EOL will act like a glue to join together the information that is already out there on the Internet. We don't have to create 1.8 million pages. What we need to do is know where the information is that's out there on the species.

How do you put it together?

The usual solution is to try and find out what the correct name is for a species and then try to get all the databases to apply the correct name. But names are always changing as a result of taxonomic research.

We adopted a different solution, which we call reconciliation. What we do is tap into the accumulated knowledge of taxonomists and draw together their understanding of all the different names that have ever been used for an organism, and extend this to cover typographical errors, vernacular names, and any name that might appear in a document that is online or a database that is online.

We've built the system so that you can

have any browsing structure in place. So if you want to change from the hierarchy provided by the Catalogue of Life [an attempt to index all known species] to that

the community we would most like to see develop a better understanding of biodiversity and how biodiversity came about.



Paddy Patterson is helping to create a web page for every known species.

provided by GenBank [a database of all published DNA sequences], you just click on the alternative classification.

How far should that flexibility extend? You previously mentioned making room for an intelligent-design perspective on the EOL?

Many bits of the EOL still need to be resolved, but one of the features of the EOL is that it has to be a highly flexible, very personalizable environment. If the EOL is to fulfil the role that we think it should fulfil, then students at high school should see something different from, for example, what a research scientist sees. Similarly, people visiting from Pakistan should be able to see the content in a language that suits them and is not culturally offensive to them.

And the same thing is true, I think, of people who hold beliefs such as intelligent design or the creationist argument. It would be, in my view, a great failing if we created an environment that they found offensive and were unwilling to enter. They are, after all, probably

That sounds like a controversial approach?

It was. Plenty of people would have me hung up by my whatever-it-was for this.

My view is that by giving the stage to this community, what they're thinking becomes open to wider scrutiny. Instead of talking in closed churches, now the public can see the extent of their thinking. I see it as something that should improve the knowledge environment in general.

You're obviously very keen on involving people in the project as widely as possible?

Some of the features we're developing will be rather like wikis or the social networking software out there. One of the things I would love to see develop early on is a 'my schoolyard' function in which kids can go outside with cell phones and take pictures of organisms and submit them to the EOL. There, the pictures are sent off to experts who verify identification. And when that is done, a little dot appears on Google Earth showing the presence of, say, a daffodil in someone's backyard.

Interview by Brendan Maher.

"Plenty of people would have me hung up by my whatever-it-was for this."

Poverty and environmental stress fuel Darfur crisis

SIR — The “true culprit” in the Darfur crisis is the National Islamic Front, according to two recent News stories (‘Darfur’s climate roots challenged’ *Nature* **447**, 1038; 2007, and ‘Darfur lake is a “mirage”’ *Nature* **448**, 394–395; 2007).

Political scientists understandably point to specific parties, leaders and military factions to account for a specific crisis such as Darfur. It is also understandable that they do not want to make excuses for perpetrators of violence. Still, as quoted, they have gone too far in the opposite direction. They seem to overlook the point that Darfur’s extreme poverty, rising population, growing water stress and desertification are all important contributors to the Darfur crisis.

The most authoritative recent study of the subject — the United Nations Environment Programme’s report *Sudan: Post-Conflict Environmental Assessment* (www.unep.org/sudan) — rightly notes that desertification, land degradation and climate change are major factors in the crisis, and appropriately comments that they are “generally contributing factors only, not the sole cause for tension”. The report also states that long-term peace in the region will become possible only if environmental and livelihood issues are resolved.

These findings are in line with a growing number of studies that find extreme poverty, falling incomes and failing rains are strong predictors of outbreaks of violence in Africa. An excellent new book *Too Poor for Peace?* (eds L. Brainard and D. Chollet, Brookings Institution Press, 2007) notes that recent research strongly favours the claim that these are the crucial drivers of conflict in less developed countries; much less solid evidence implicates political repression.

It is important that considerations of ecology, climate and extreme poverty are integrated with political analyses of conflict. This is especially the case as demographic pressures hit against resource limitations and climate change in the Horn of Africa and elsewhere.

Jeffrey D. Sachs

The Earth Institute at Columbia University, 535 West 116th St, New York, New York 10027, USA

A paper should appear with all the information it needs

SIR — Until the past few years, both *Nature* and *Science* confined their articles and letters to a rather small number of words. This was both good and bad; good in that the articles were short and to the point; bad in that it eliminated studies that were complex. I first

thought that the Supplementary Information sections were a great idea. Here was a way to place at the readers’ disposal important data (tables or figures) that were necessary background to the work, but not necessary to the reading and understanding of the paper.

However, some recent articles refute my thinking. One or two have contained tens of pages of this supplementary material, essential to the reading and understanding of the article. Ten pages of Supplementary Information are not unusual, and the average for *Nature* is about five pages.

I suggest either that you either publish hard-copy papers whole and integrated in a long form, or publish them whole and integrated on the web, as you now do with Methods sections.

Larry Benson

Chief Arid Regions Climate Project, USGS Suite E127, 3215 Marine St, Boulder, Colorado 80303-1066, USA

Comments are welcomed at Nautilus:
http://blogs.nature.com/nautilus/2007/09/papers_should_not_need_supp.html

Diverse approaches useful for microbicide trials

SIR — Your News story ‘HIV trial doomed by design, say critics’ (*Nature* **448**, 110–111; 2007), portrays a biased and excessively pessimistic view of the Centre for the AIDS Programme of Research in South Africa (CAPRISA) 004 trial. As the first microbicide trial to be initiated from, and partially funded by, South Africa, the CAPRISA 004 trial promises to address some fundamental questions about the safety and effectiveness of tenofovir as a microbicide gel.

More published data exist on tenofovir for HIV prevention than for any other microbicide; thirteen monkey and four human tenofovir studies were considered when designing the CAPRISA 004 trial.

In the rural South African community where CAPRISA runs a large AIDS treatment programme, women with migrant partners account for most new HIV infections. The CAPRISA 004 trial and its episodic dosing strategy of two doses in 24 hours, one before and one after intercourse, was designed to maximize product adherence and hence the presence of tenofovir at exposure, especially for women with migrant partners.

This dosing strategy was selected after extensive consultation with international scientific experts, including those knowledgeable about microbicides, tenofovir and clinical-trial design. Just as importantly, it followed detailed consultation with the communities involved. Moreover, to assess the effectiveness of this strategy, extensive data will be collected on gel use, timing of gel

use in relation to intercourse, adherence to the dosing regimen and tenofovir levels.

Differences of opinion and interpretation of data are important in science, including research on women-initiated HIV prevention, as you note in your Editorial ‘Transmission lines’ (*Nature* **448**, 225–226; 2007). But in the absence of a marker that can show whether a particular agent, dosing strategy or trial design will lead to protection against HIV, only the trial results will reveal the truth.

Salim S. Abdool Karim, Quarraisha Abdool Karim
CAPRISA, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa

DAMA sheds light on dark-matter particles

SIR — We were disconcerted to see that the experiment XENON10 was the main focus of your News Feature ‘Welcome to the dark side’ on searches for particle dark matter (*Nature* **448**, 240–245; 2007), even though it is still in its infancy. The well-established DAMA/NaI experiment, on the other hand, was confined to a box titled ‘Contested results’.

Experiments searching for dark-matter particles must be able to discriminate between a dark-matter signal and background signals created by other particles. Because the number of dark-matter particles hitting Earth is expected to vary during the course of a year, one strategy is to look for an annual variation (with many specific features) in the frequency of peculiar events registered by a detector. Running successfully for seven years, the DAMA/NaI experiment, with its 100-kilogram sodium iodide target, has detected such a peculiar annual variation, the properties of which meet those expected of a dark-matter signal with a confidence level of more than six standard deviations.

It has been proved quantitatively that no effect could produce this signal apart from dark-matter particles. Also, the DAMA effect is not contradicted by the results of other experiments because these experiments take a different approach, and so are unable to investigate the same effect.

The DAMA effect has been proved to be compatible with the most popular dark-matter candidate: the neutralino. Moreover, the DAMA experiment is sensitive to physical scenarios and candidates other than the neutralino — to which other experiments are either poorly sensitive or even blind.

Rita Bernabei* (on behalf of the DAMA Collaboration), Alessandro Bottino†

*Department of Physics, University of Rome Tor Vergata, and INFN Rome Tor Vergata, via della Ricerca Scientifica 1, I-00133 Rome, Italy

†Department of Theoretical Physics, University of Torino, and INFN Torino, via P. Giuria 1, I-10125 Torino, Italy

BOOKS & ARTS

An act of communal imagination

A Disappearing Number, a play exploring the partnership between mathematicians G. H. Hardy and Srinivasa Ramanujan, opens in London this week. It is the latest of several astonishing works devised by leading international theatre company Complicite, marbled with science and technology. Artistic director **Simon McBurney** tells *Nature* about the results of his most recent round of collaboration and experimentation.

What drew you to the tale of Hardy, the maths professor, and Ramanujan, his self-taught Indian protégé?

I grew up in Cambridge, surrounded by academics' stories of brilliant people, including these two. Then, around 1997 I was in Toronto talking to the writer Michael Ondaatje about creativity — in a bowling alley. He said, "The best book I know about creativity is Hardy's *The Mathematician's Apology*", and he gave me a copy.

It haunted me. I started to read about how, in 1913, Hardy got a letter from Ramanujan so interesting that Hardy brought him to study at Cambridge University. The pioneering work they did together meant that the frail Ramanujan was made a fellow of the Royal Society before he returned to Tamil Nadu in India. By 1920 he was dead.

After I had made *Mnemonic* — our play on the discovery of Ötzi, the 5,000-year-old 'iceman' — the book just kept nagging at me. So I did some research with the mathematician and actress Victoria Gould.

We discovered that this almost mythological story of somebody finding somebody else at the beginning of the twentieth century is important on many levels: for the scientific and mathematical ideas that are still being used; for the image of collaboration; for the notion of misunderstanding or accepting another culture, which is still present.

Composer Nitin Sawhney has created tabla lines based on Ramanujan's work, and mathematician Marcus du Sautoy inducted Complicite into primes and hypergeometric series. Even the president of the Royal Society, Martin Rees, advised. What have you discovered through this collaboration?

I learnt that mathematics is a relay race. This provides an important image of human continuity in these egotistical times. In biology or physics or chemistry, people might have an idea that is then completely refuted, but Euclid's proof of the infinity of primes or Pythagoras's proof of the irrationality of the square-root of 2 are immutable.



G. H. Hardy (played by David Annen) saw mathematicians as pattern-makers, like poets or painters.

Likewise, I am constantly in a relay race. Only when somebody gives me an idea can I transform it into something else. If I sit by myself in an office to write down ideas, they vanish. The moment I'm in contact with other people, they seem to flow.

Hardy felt that a mathematician, like a painter or a poet, is a maker of patterns. A director is also a maker of patterns. Theatre is what the audience makes of those patterns. It is an act of communal imagination, of collaboration. It is no coincidence that theatre audiences are about the size of early human communities. Theatre touches on people's fundamental need to connect.

Like a lab, Complicite creates through years of experiment and iteration. How?

I am completely remaking *A Disappearing Number* for the Barbican Theatre in London from the show that toured Europe earlier this year. Each time you work at something again,

you try to make it a little more accurate, to take it another step further, to make it more compelling, more comprehensible.

There is a great freedom in saying "let's just throw this away". At the same time, when you discard ideas, you have to be careful not to take away what was instinctive and intuitive. You can be left with something much too simplistic. This brings us back to Ramanujan and Hardy. Ramanujan lived with an enormous amount of mystery and, in mathematical terms, roughness. He was constantly guessing and approximating, nonetheless coming up with extraordinary ideas. Hardy was a great deal more disciplined in the way that he created proofs. In the end, some say the mathematics of Ramanujan is much greater than that of Hardy.

***The Elephant Vanishes*, your 2003 show, investigated the wave-particle duality**

of light. In 1999, *Mnemonic* juggled archaeology and neurochemistry. Might audiences be more afraid of maths than any of these topics?

We're addressing that. The play begins with the explanation of the functional equation of the Riemann zeta function — to do with the distribution of primes — and that is as difficult as it gets. Even if the audience doesn't understand the mathematics, they start to get a sense that it can be beautiful, simply for its elegance and economy. Great ideas themselves are touching, in the same way that a human story is touching.

Hardy was the only person who could recognize how incredible Ramanujan's work was, because he could appreciate something enormously, even if it was plain wrong. The pattern of how or why it was wrong

fascinated him — as we might be enchanted by any other work of art.

What idea would you like to play with next?

Consciousness. Theatre must always be interested in what we don't know. We still can't explain consciousness — whatever Daniel Dennett might say. Putting things we don't understand on stage is a process of trying to learn about them: you communicate something of what you're learning, and perhaps take the audience on the journey you are in. ■

Interview by Sara Abdulla, *Nature's* chief commissioning editor.

***A Disappearing Number* runs at the Barbican Theatre in London from 5 September to 6 October (www.complicite.org).**

tradition was revived in the eighteenth century by William Hogarth and Francisco de Goya.

Until Darwin came along, such cross-attribution never shook the deeply held belief that humans, with their capacity for abstraction, were cleanly distinct from animals, with their inability to rationalize their feelings or control their instincts, appetites and passions. In the seventeenth century, René Descartes argued that animals were but machines without souls, incapable of experiencing the finer emotions that elevate humans.

That concept was dangerously extended in the eighteenth century by Julien Offray de La Mettrie, who described humans as 'perfect machines'. He was not putting the human soul in doubt — but European philosophy was certainly moving along a path that spelt trouble for God. Automata became popular in La Mettrie's time. These 'living' or 'philosophical' machines could accurately reproduce particular behaviours of animals and so demonstrate the redundancy of the soul. The 'digesting duck' of Jacques de Vaucanson, for example, could eat grain and apparently expel the digested waste from its anus (a mechanical achievement not diminished by the fact that the chemical laboratory claimed for the stomach was later found to be absent). The greater challenge for philosophers was Wolfgang von Kempelen's 1770 chess-playing automaton, 'the Turk'. It seemed to have the ability to reason, a defining characteristic of 'humanness'.

The boundaries between what is human and what is animal became increasingly blurred in that century, with the rise to fame of some feral children raised alone in the forest, who had a limited ability to learn to speak or behave in other ways considered to be human. Travelling circuses and freak shows were popular. They displayed animals trained to do 'human'

The animal in us

The Human Animal in Western Art and Science

by Martin Kemp

University of Chicago Press: 2007. 320 pp. \$40

Alison Abbott

On waking, Henry Jekyll stared with horror at the metamorphosis of his hand, normally "professional in shape and size... large, firm, white and comely". Jekyll's experiment to separate the human and animal sides of himself had been all too successful. He noted further: "The hand which I now saw... lying half shut on the bed-clothes was lean, corded, knuckly, of a dusky pallor and thickly shaded with a smart growth of hair. It was the hand of Edward Hyde."

Thus Martin Kemp ends his treatise *The Human Animal in Western Art and Science* with this apposite quote from Robert Louis Stevenson's 1886 novel. It epitomizes the dilemma that has fascinated us for millennia. How much of the animal is there within us? Conversely, how much is human in animals?

Kemp answers these questions. Science, from Darwin to the latest neuroscience and genomics, has shown that there is no sharp animal-human divide, only a sliding scale. And in guiding us to this conclusion, Kemp's six chapters deviate through an amusing and erudite visual history, drawing from art, philosophy, literature, film and other cultural media.

We humans have always had a tendency to anthropomorphize, and no amount of science will erase our pleasure in imagining the lion as fierce but noble and generous, the snake as cold and deceitful. We also instinctively assign animal labels to our moods and attributes, a tendency frequently exploited over the centuries as a literary device. In the fables of Jean de La Fontaine (1621–95), so beautifully illustrated in the eighteenth century by Jean-Baptiste Oudry,

animals endowed with specific human failings (and speech) enact tales of eternally relevant morality. And when negotiations between the head pig and the farmer in George Orwell's 1945 *Animal Farm* become ugly, the other creatures "looked from pig to man, and from man to pig again; but already it was impossible to say which was which".

In classical times, the theory of physiognomics attempted to provide a rational framework for all this. Those with the broad brow and square face of a lion might be expected to share the lion's perceived nature, for example. Artists have often used such physiognomics to inform their portraits — Albrecht Dürer for sure, and Rembrandt, ventures Kemp, and the



This sketch by Charles Le Brun of a man with a beak-like nose plays on the similarities between animals and humans.

MUSÉE DU LOUVRE, PARIS/PHOTO RMN/M. BELOT

tricks, such as the learned pig, which could spell out words. They also displayed unfortunate humans, such as microcephalics and hairy women, who were branded as 'freaks' for showing animal characteristics.

The boundaries came close to merging in the nineteenth century. The widespread study of primate anatomy and behaviour made the similarities between humans and apes hard to dismiss. Satirists saw this as an opportunity: for example, one hairy creation, Sir Oran Haut-Ton, bought a baronetcy and entered parliament — where, his creator reported, little distinguished his behaviour from that of his fellows aside from his muteness. Darwin's influence on our mental image of primitive man is unmistakable. In Renaissance pictures, he is clearly *Homo sapiens*, albeit extremely unkempt. Post-Darwin paintings endow primitive man with decidedly simian characteristics.

Kemp's book is beautifully illustrated and written, but the breadth of its discussion is packed into a form that may be too abbreviated. It is not a light read, and sometimes one struggles to follow the thread of the narrative. Then again, it is not supposed to be a linear plot. How else could all this knowledge be shoe-horned into a coffee-table format?

Also packed in are discussions of the science of phrenology, as developed by Franz Joseph Gall in the first half of the nineteenth century, and the fashion for quantification of primitiveness through measurement of jaw angle — the ideal reference being Apollo Belvedere (who was decidedly caucasian) from antiquity.

The great French nineteenth-century neuroscientist Paul Broca also made the error of assuming that caucasians must provide the baseline. However, he was forced to abandon his theory that intelligence was related to brain size when facts intervened — his own measurements showed that the brains of some non-caucasian races were in fact larger. His contemporary Cesare Lombroso, who launched the field of criminal anthropology, even tried to match skulls with particular shapes, and brains with particular features, to show that those with criminal tendencies had the physical attributes of the throwback — or of other animals such as apes, or even rats.

With hindsight, it is easy to scoff at these past theories. Instead, Kemp acknowledges most of them as appropriately reflecting the state of knowledge at the time. Gall's phrenology, he notes, had its own internal logic and was based on careful, systematic measurements.

Robert Louis Stevenson was also of his scientific time. His Mr Hyde is a Lombroso-style atavistic regression, exemplifying the thin veneer of civilization that only just manages to repress our natural animalistic tendencies. Dr Jekyll's dilemma resonates as strongly in today's sophisticated times — the adjective 'Jekyll-and-Hyde' has made its way into the *Oxford English Dictionary*. ■

Alison Abbott is *Nature's* Science in Culture editor and senior European correspondent.

A question of truth

A Certain Ambiguity: A Mathematical Novel

by Gaurav Suri and Hartosh Singh Bal

Princeton University Press: 2007. 304 pp.

£16.95, \$27.95

Katherine Körner

When Euclid laid out his axioms of geometry — five statements on the properties of points, lines, circles and angles that he believed to be self-evident — he sought to demonstrate further truths by building logically on these supposedly indisputable foundations. *A Certain Ambiguity* uses the structure of a novel to explore the nature of the beauty inherent in such mathematical arguments, and of the truth to which they lead.

The authors' commitment to wowing their characters as much as their readers is one of the book's strongest points. Gaurav Suri and Hartosh Singh Bal draw on some of the most elegant results in the ancient Greek arsenal — proofs that there are infinitely many prime numbers, for instance, and that the square root of 2 cannot be written as a fraction. The book's characters and readers are led through deep (but accessible) reasoning before being challenged with questions about what makes these arguments valid, and on what this validity depends.

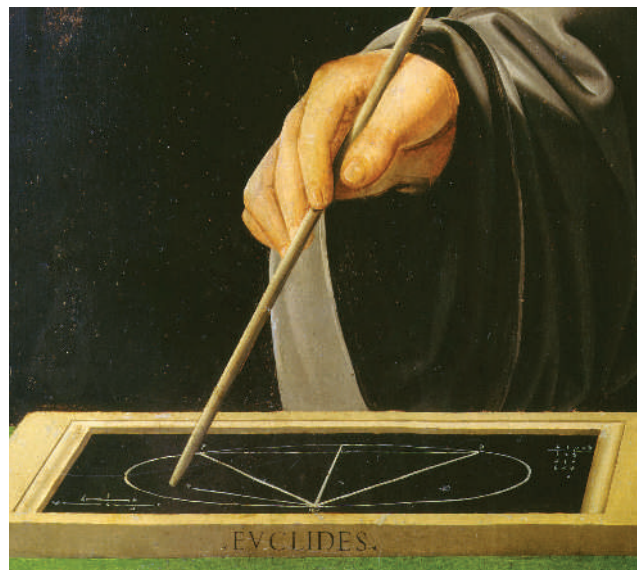
The book is at its most successful, and the story at its most engaging, when the authors allow their natural exuberance to carry them away. The joy that accompanies their early excursions into proof will buoy up a floundering newcomer, while allowing old hands to recall the thrill of meeting the ideas for the first time.

Through a series of court transcripts rationed out over the course of the book, the narrator's grandfather, under arrest for blasphemy in 1919, talks his judge, his grandson and us through the components of logical argument and his belief in the truth derived from it. Only the judge is allowed to answer back, and the discussions that ensue are by far the most gripping and entertaining aspect of the book. Together, judge and grandfather journey from their delight at finding a way through simple proofs, to awe at the world this reveals, all the way to spiritual torment when the truth of Euclid's axioms is brought into question. In one of the most profound sections, the authors offer a lucid introduction to the rejection of one

of these axioms by noneuclidean geometry.

We also follow the less-stimulating conversations of the narrator with his undergraduate friends as they, too, grapple with the nature of truth. Although their mathematical focus on our knowledge of infinity is fascinating and well presented, the discussions themselves are common to those of many students in the pub at closing time and do not particularly enlighten the reader or add much of interest. Each of these characters seems to have had his or her list of personality traits allocated at random, and it was galling to see 'love interest' on the list of the only important female character flagged up from early on.

The book's main weakness is that these same characters are tedious. The grandfather's story



Euclid's axioms are used to demonstrate other geometric truths.

is compelling, and the combination of socratic dialogue and very human digressions throughout should interest those readers less inclined to mathematical textbooks. But — just as the characters' excitement in mathematics encourages the reader's — when it is hard to warm to the characters, it is hard to care about the topics they choose to discuss.

In any introduction to a subject, there will be areas the specialist might wish to see covered in more depth or from a different angle. The interested layperson, however, should take from *A Certain Ambiguity* a good grounding and a new curiosity that will help them tackle the books referenced in the bibliography. Mathematicians should enjoy the infectious enthusiasm with which the discipline is presented, and it could even spur them into finding new ways to share their own passion. ■

Katherine Körner is a graduate student in the Department of Mathematics, Harvard University, 1 Oxford Street, Cambridge, Massachusetts 02138, USA.

A. DAGLI ORTI/MUSEO DI CAPODIMONTE, NAPLES/ART ARCHIVE

NEWS & VIEWS

HIV/AIDS

Allied responses

John R. Mascola

The main function of neutralizing antibodies is to block viral entry into host cells. But, for maximal protection against HIV, such antibodies must call upon other elements of the immune system to help with the job.

More than two decades into the AIDS pandemic, and despite intense research on the human immunodeficiency virus (HIV), we still do not have a vaccine to prevent HIV infection. One possibility is to trigger the components of adaptive (specific) immunity¹ — T cells and B-cell-derived antibodies. These immune mediators can function near mucosal surfaces, which are in contact with the external environment and are therefore sites of viral entry and replication. For complete protection against infection, however, a vaccine probably needs to elicit neutralizing antibodies to block viral entry into host cells. How do neutralizing antibodies against the two HIV species — HIV-2 and the more virulent HIV-1 — function? On page 101 of this issue, Hessel and colleagues² report on the mechanism through which one of the best-known neutralizing antibodies to HIV-1 functions *in vivo*.

In infected patients, the production of neutralizing antibodies that would target most circulating strains of HIV-1 is uncommon, and

triggering such antibodies through vaccination has proved difficult³. During the past 15 years, only a handful of broadly neutralizing monoclonal antibodies (antibodies that recognize a single marker on its target) have been isolated from infected patients³. One of these, known as b12, was isolated in 1992 and has been widely studied since⁴.

Previous work^{4,5} had shown that the b12 antibody neutralizes HIV-1 by attaching to a receptor-binding site on the virus's envelope glycoprotein, thereby blocking viral access to its primary receptor, the CD4 molecule, which occurs on the surface of T cells (CD4⁺ T cells). And when b12 is intravenously infused into monkeys, it prevents infection of mucosal tissues with a chimaeric simian/human immunodeficiency virus (SHIV)⁶.

Generally, the level of protection elicited by an anti-HIV antibody relates to its neutralizing potential, which is measured by its ability to prevent infection of CD4⁺ T cells *in vitro*. Thus, the work on b12 and several other neutralizing

monoclonal antibodies^{6–9} suggests that, to be effective, an HIV-1 vaccine should induce a potent neutralizing-antibody response. However, virus neutralization is not the only function of such antibodies.

Antibodies such as b12 consist of two identical heavy chains and two identical light chains. The carboxy-terminal regions of the heavy chains contain the Fc region that binds both to proteins of the complement system (involved in killing pathogens) and to Fc receptors on immune cells such as neutrophils, mononuclear phagocytes and natural killer cells. Phagocytic cells then engulf antibody-coated viruses, whereas natural killer cells destroy virally infected cells coated with antibodies to viral antigens by a process known as antibody-dependent cellular cytotoxicity (ADCC). For several viral diseases, these Fc-mediated antibody functions are thought to be important¹⁰, but their contribution to protection against HIV-1 has not been well documented.

Hessel and colleagues² set out to study the

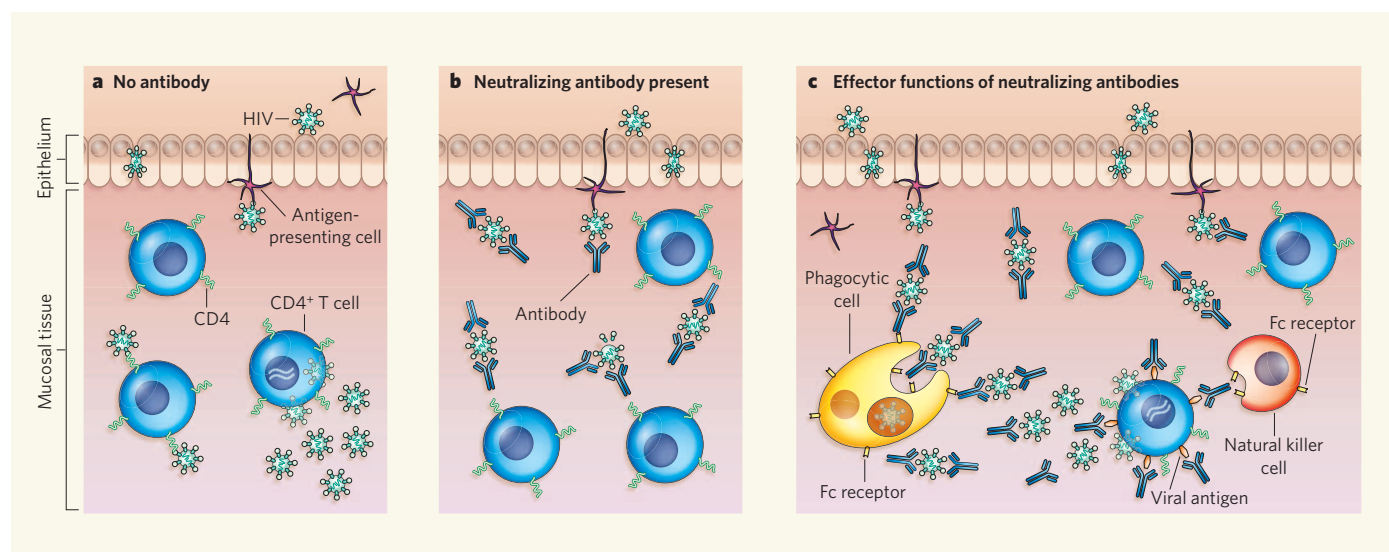


Figure 1 | Antibody activities in mucosal tissues. HIV-1 can reach mucosal tissues by means of a breach in the epithelial cell layer, which is in contact with the external environment, or by interacting with antigen-presenting cells. **a**, In the absence of antibodies, the first few CD4⁺ T cells infected produce additional infectious virus particles, leading to their local spread. **b**, If present, neutralizing antibodies react with free virus particles, blocking their infection of CD4⁺ T cells. **c**, Hessel *et al.*² show that protective antibodies

are most effective when they can deploy additional elements of the immune system that are activated by the Fc region of the antibody molecule. For example, the engulfing of antibody-coated viral particles by phagocytic cells may contribute to viral clearance. Moreover, the few T cells that become infected and express viral antigens could be immediately coated with antibodies and then eradicated by the process of antibody-dependent cellular cytotoxicity (ADCC), which is mediated by natural killer cells.

mechanism of b12 action in a biologically relevant animal model. They inoculated monkeys with SHIV in order to study the protective effect of normal b12, as well as of two mutated versions, K322A and L234A–L235A (LALA). These differ from normal b12 by only one and two amino acids, respectively. The ability of the K322A variant to bind to complement proteins is compromised, whereas the LALA variant cannot bind to either complement proteins or Fc receptors². *In vitro*, all three antibodies have similar virus-neutralizing potencies.

To test the potency of these antibodies *in vivo*, the authors intravenously infused female monkeys with individual antibody variants, then introduced SHIV vaginally. The normal b12 antibody, as well as the complement-defective K322A mutant, protected eight out of nine animals from SHIV infection, whereas the Fc-receptor-defective LALA variant protected only five out of nine.

There are two ways to look at these data. On the one hand, it can be argued that even without complement- or Fc-mediated effector functions, the b12 antibody can protect more than half of the animals. Such a conclusion substantiates our current belief that antibody-mediated protection against HIV is conferred mainly through direct neutralization of the virus. On the other hand, the work of Hessel *et al.* shows that Fc-mediated effector functions are necessary for the antibody to have its full protective effect. The group's findings indicate that clearance of antibody-coated viral particles by phagocytic cells, or the killing of virally infected cells by ADCC, could be contributing to maximal protection against HIV infection. Although the authors do not show how the binding of b12 to Fc receptors might contribute to protection *in vivo*, they provide *in vitro* data showing that normal b12 and K322A, but not the LALA variant, mediates ADCC. In addition, ADCC has been associated with antibody-mediated protection in monkeys infected with simian immunodeficiency virus^{11,12}.

So a picture emerges (Fig. 1) in which either passive infusion of antibodies or active immunization results in circulating antibodies in mucosal tissues. Subsequently, on inoculation of a mucosal surface with HIV-1, neutralizing antibodies react with most of the free virus particles to prevent the infection of CD4⁺ T cells. Some free viruses are also coated by antibodies and destroyed by phagocytic cells. But a few CD4⁺ T cells become infected and express viral antigens on their surface. Before these cells can initiate further rounds of productive infection, however, antibodies tag them for eradication by Fc-receptor-bearing natural killer cells.

That antibodies exert their protective effect by mobilizing an allied army of immune cells may have important implications for the design of HIV vaccines. The robustness of vaccine-mediated antibody responses are commonly measured by quantifying virus neutralization, but such analysis does not measure Fc-mediated effector functions of the antibody.

From Hessel and colleagues' results, these commonly used neutralization assays might be unable to fully predict a vaccine's efficacy. Moreover, some might argue that the authors' findings suggest that an effective vaccine could elicit non-neutralizing antibodies that mediate ADCC, but do not block infection of target cells.

I would urge caution. Clearly, there is much we don't know about how protective antibodies work and the level of antibodies required for maximal protection. Nonetheless, results of previous animal and human studies^{7–9,13} show that non-neutralizing antibodies alone have little or no protective effect. For example, a large clinical study¹³ found that a vaccine consisting of a soluble form of the HIV envelope glycoprotein gp120 generates high levels of non-neutralizing antibodies that have no effect on HIV-1 infection.

As existing HIV-1 vaccines do not elicit potent neutralizing antibodies, the focus of vaccine researchers and funding agencies on identifying better vaccine antigens seems completely appropriate. At the same time, we should be aware of the complexity of the

immune responses mediated by secretory antibodies. As Hessel and colleagues' findings² indicate, antibodies against HIV-1 are likely to exert a protective effect through more than one mechanism. So continued research should seek to elucidate and measure alternative biological functions of protective antibodies.

John R. Mascola is at the Vaccine Research Center, NIAID, NIH, 40 Convent Drive, MSC 3015, Bethesda, Maryland 20892, USA.
e-mail: jmascola@nih.gov

1. Johnston, M. I. & Fauci, A. S. *N. Engl. J. Med.* **356**, 2073–2081 (2007).
2. Hessel, A. J. *et al. Nature* **449**, 101–104 (2007).
3. Haynes, B. F. & Montefiori, D. C. *Expert Rev. Vaccines* **5**, 579–595 (2006).
4. Burton, D. R. *et al. Science* **266**, 1024–1027 (1994).
5. Zhou, T. *et al. Nature* **445**, 732–737 (2007).
6. Parren, P. W. *et al. J. Virol.* **75**, 8340–8347 (2001).
7. Mascola, J. R. *et al. Nature Med.* **6**, 207–210 (2000).
8. Shibata, R. *et al. Nature Med.* **5**, 204–210 (1999).
9. Mascola, J. R. *Curr. Mol. Med.* **3**, 209–216 (2003).
10. Reading, S. A. & Dimmock, N. J. *Arch. Virol.* **152**, 1047–1059 (2007).
11. Gómez-Román, V. R. *et al. J. Immunol.* **174**, 2185–2189 (2005).
12. Binley, J. M. *et al. Virology* **270**, 237–249 (2000).
13. Graham, B. S. & Mascola, J. R. *J. Infect. Dis.* **191**, 647–649 (2005).

SOLAR SYSTEM

Lethal billiards

Philippe Claeys and Steven Goderis

A huge collision in the asteroid belt 160 million years ago sent fragments bagatelling around the inner Solar System. One piece might have caused the mass extinction that wiped out the dinosaurs 65 million years ago.

What are the chances of the sky falling on our heads? If it's an asteroid hitting Earth that we're talking about, we can't be too sure. Not only do estimates of the current terrestrial meteorite impact rate differ by a factor of five to ten, depending on the approximations used¹, but we don't know for certain whether that rate has remained constant or has varied throughout geological time.

A current theory proposes that the impact rate has increased during the past 100 million years or so. On page 48 of this issue, Bottke *et al.*² present an intriguing explanation for why this might be, invoking errant fragments from a powerful ancient collision in the asteroid belt between Mars and Jupiter.

Clusters of impact craters and layers of material ejected in meteorite impacts, as well as higher levels of extraterrestrial material in some sedimentary rocks, seem to indicate that, during several glacial periods, the Earth–Moon system has suffered abnormally high rates of bombardment. The late Miocene epoch around 8 million years ago, for example, was marked by an increased flux of interplanetary dust particles between 1 µm and 1 mm across, which might have been produced by a collision within the

asteroid belt³. An asteroid or comet shower has similarly been put forward to explain the higher dust-particle flux in the late Eocene around 35 million years ago, an event that seems to be coupled with an unusually high concentration of impact craters^{4,5}. These include the two largest craters in recent geological history, Popigai in Siberia (100 km in diameter) and Chesapeake Bay off the Maryland coast (around 85 km in diameter).

And we can go even farther back in recording periods of heavy bombardment. The abundant micrometeorites in the 480-million-year-old Ordovician limestones of southern Sweden most probably reached Earth after a significant disruption had occurred in the asteroid belt⁶. Several impact craters also seem to cluster around this age, although here the geological record is rather poor. Farther back still, recognized ejecta layers are concentrated in two time windows between 2.65 billion and 2.5 billion years ago and 3.47 billion and 3.24 billion years ago⁷. Finally, the most dramatic series of events is undoubtedly the Late Heavy Bombardment of 3.8 billion years ago, the occurrence of which is inferred from the lunar cratering record⁸. Although its traces have been erased

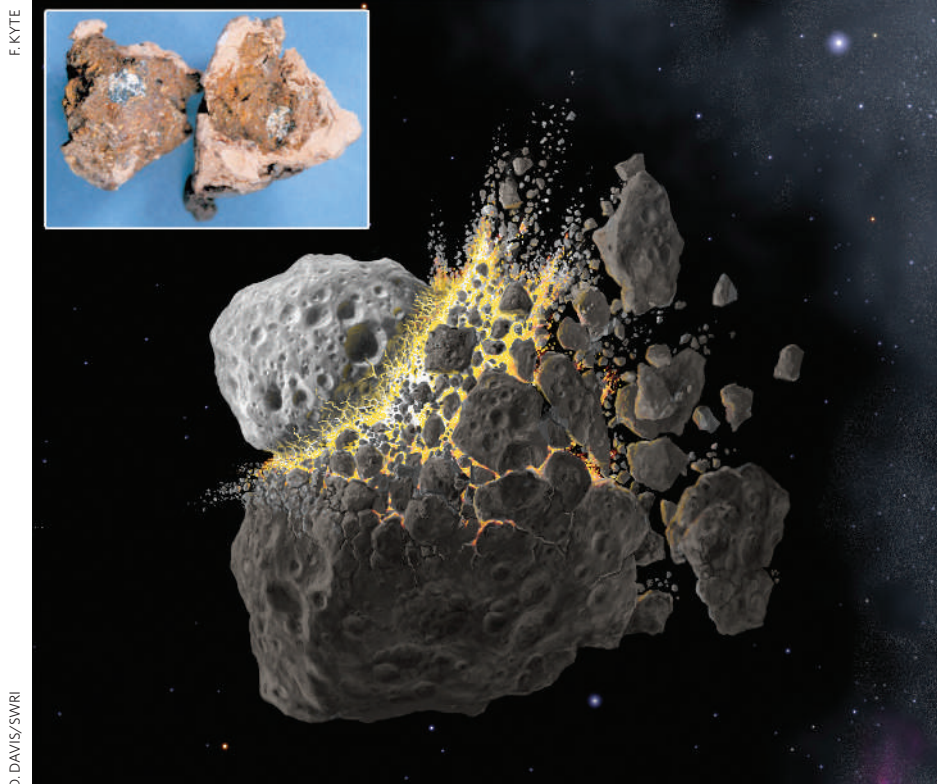


Figure 1 | Breaking up is all too easy. Bottke and colleagues² propose that the collisional break-up of an asteroid 160 million years ago was the ultimate cause of the mass extinction on Earth at the Cretaceous–Tertiary boundary some 95 million years later. This occurred when one particularly large fragment, the Chicxulub impactor, hit Earth. The inset shows an inferred remnant of this meteorite, retrieved from clays in the northern Pacific dating to the same time as that event¹¹.

by geological activity on Earth, extrapolation of the lunar data indicates⁹ the formation of up to 22,000 terrestrial craters with a diameter of more than 20 km. This catastrophic bombardment probably resulted from colliding asteroids disturbed by changes in the orbits of the giant gas planets¹⁰.

Bottke *et al.*² have discovered the remnants of another huge collision hidden in the inner region of the main asteroid belt. These comprise the Baptistina asteroid family (BAF), a class of variously sized objects of similar composition and orbital geometry, typified by the 40-km-diameter asteroid known as (298) Baptistina. The authors use a computer simulation to track the orbits of these fragments back to the moment of their formation, and find that the collision must have taken place about 160 million years ago. The best fit to the data is given by an object of 60-km diameter colliding almost vertically with a 170-km-diameter body. This collision, at a velocity of 3 km s⁻¹, generated more than 1,000 large bodies greater than 1 km in diameter (Fig. 1).

The authors' dynamic modelling indicates that the Baptistina disruption took place in a region of the asteroid belt where the gravitational influence of Mars and Jupiter would have caused a number of large objects to be perturbed into Earth-crossing orbits. The collision of just a small fraction of these fragments with Earth would account for the increased

cratering rate during the Cretaceous period (from around 145 million to 65 million years ago) and the early Cenozoic era that followed. The fragments would in fact have formed an asteroid shower in the inner Solar System lasting some 100 million years.

According to spectroscopic measurements, the composition of (298) Baptistina is similar to that of primitive 'carbonaceous chondrite' meteorites, a rather infrequent type of meteorite that contains unusually large amounts of water and organic compounds. This is front-page news, as a 10-km-sized carbonaceous chondrite is most probably the projectile that formed the Chicxulub crater on the Yucatán peninsula in Mexico^{11,12} (Fig. 1, inset). This impact almost certainly triggered what is known as the K/T mass extinction, including the demise of the dinosaurs, at the boundary between the Cretaceous and Tertiary periods 65 million years ago.

By comparing the impact rate of BAF near-Earth objects with the background occurrence of objects with a similar composition, Bottke and colleagues² estimate that there is a 90% chance that the Chicxulub projectile came from the BAF. Other craters formed during the same period on the inner bodies of the Solar System could share the same origin: one candidate is Tycho on the Moon, which was formed 109 million years ago at what would have been the climax of the BAF asteroid shower.

This hypothesis is nothing if not provocative. It implies that a significant number of terrestrial craters that formed during the past 160 million years resulted from carbonaceous chondrite projectiles. The concentrations of platinum-group elements and the chromium-isotopic signatures measured in rocks melted in meteorite impacts provide precise information on the projectile type, and can be used to distinguish carbonaceous chondrites from other types of meteorite. Of the eight craters on Earth that are larger than 1 km in diameter (implying a projectile 50 m or more across) and less than 200 million years old for which the projectile composition has been unequivocally identified, Chicxulub stands out as an anomaly — it is the only one formed by a carbonaceous chondrite¹³. So did the other, smaller BAF objects miss Earth, or is this apparent anomaly due to the fact that projectile type is well characterized for relatively few craters?

That question must remain unanswered for now. Nevertheless, unless a rogue comet came from the outer edge of the Solar System (a rather unlikely event), the BAF remains a likely source for the Chicxulub projectile. It is a poignant thought that the Baptistina collision some 160 million years ago sealed the fate of the late-Cretaceous dinosaurs well before most of them had even evolved.

The most important point raised by Bottke and colleagues' discovery² of the Baptistina asteroid family is how severe the repercussions of cataclysmic collisions in the asteroid belt can be for the Earth–Moon system. The terrestrial impact record needs to be scrutinized more closely to identify and understand these periods of more intense bombardment, and to link them to the huge and dangerous game of billiards continuously being played out between the orbits of Mars and Jupiter.

Philippe Claeys and Steven Goderis are in the Department of Geology, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. e-mail: phclaeys@vub.ac.be

- French, B. M. *Traces of Catastrophe: A Handbook of Shock-Metamorphic Effects in Terrestrial Meteorite Impact Structures* (Lunar Planet. Inst., Houston, TX, 1998).
- Bottke, W. F., Vokrouhlický, D. & Nesvorný, D. *Nature* **449**, 48–53 (2007).
- Farley, K. A., Vokrouhlický, D., Bottke, W. F. & Nesvorný, D. *Nature* **439**, 295–297 (2006).
- Farley, K. A., Montanari, A., Shoemaker, E. M. & Shoemaker, C. S. *Science* **280**, 1250–1253 (1998).
- Tagle, R. & Claeys, P. *Science* **305**, 492 (2004).
- Heck, P. R., Schmitz, B., Baur, H., Halliday, A. N. & Wieler, R. *Nature* **430**, 323–325 (2004).
- Simonson, B. M. & Glass, B. P. *Annu. Rev. Earth Planet. Sci.* **32**, 329–361 (2004).
- Ryder, G., Koeberl, C. & Mojzsis, S. J. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 475–492 (Univ. Arizona Press, Tucson, 2000).
- Kring, D. A. & Cohen, B. A. *J. Geophys. Res.* **107**, 10.1029/2001JE001529 (2002).
- Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. *Nature* **435**, 466–469 (2005).
- Kyte, F. T. *Nature* **396**, 237–239 (1998).
- Shukolyukov, A. & Lugmair, G. W. *Science* **282**, 927–929 (1998).
- Tagle, R., Goderis, S. & Claeys, P. *Lunar and Planetary Science Conference 2007*, League City, TX, abstr. 2216.

EARTHQUAKES

Burma's fault

The great Sumatra–Andaman earthquake of 26 December 2004, and the tsunami that it triggered, also shook the geological community. Much scientific effort has since focused on the possibility of further calamitous events in the Bay of Bengal, and especially on understanding the southern stretch of the fault that was responsible for the earthquake, and which lies to the west of the island of Sumatra.

But the fault's northernmost limit, which extends along the coast of Myanmar (Burma) to Chittagong in Bangladesh and faces the densely populated Ganges delta (pictured), has received relatively little attention. Phil Cummins' conclusion that an active zone still exists off Myanmar, as he reports on

page 75 of this issue, thus makes for disquieting reading (P. R. Cummins *Nature* **449**, 75–78; 2007).

The Sumatra–Andaman earthquake was triggered when the Indian tectonic plate was thrust violently under the southeast Asian plate off the northwestern coast of Sumatra. Although this subduction zone was known to extend farther north through the Andaman Islands, its location and nature in the Myanmar region were less clear. It had been thought that subduction was no longer active in this area, and that the plate boundary extended, not under the sea, but on the land through Myanmar.

Cummins bases his alternative hypothesis on previous geological studies and recent geodetic

measurements, as well as on historical accounts detailing the effects of past earthquakes.

He couples these observations with the fact that the floor of the Bay of Bengal has a thick layer of sediments, up to 20 kilometres deep, fanning out from the mouths of the Ganges and Brahmaputra rivers. This 'Bengal fan' insulates the underlying rock, creating thermal conditions more suited to generating earthquakes. Taken together, the implication is that a long, submarine subduction zone stretches for some 900 kilometres from the northern Andamans to the west of Chittagong.

The existence of such a fault requires a thorough re-evaluation of the potential for deadly



tsunamis in the northern Bay of Bengal. Cummins' simulation of a large earthquake off the coast of Myanmar and the resultant tsunami shows the devastating effect it could have, and underscores the need for further study and monitoring of rock deformation in this region.

Ninad Bondre

Ninad Bondre is an associate editor at *Nature Geoscience*.

ZOOLOGY

Twice bitten

Mark W. Westneat

The toothy visage of a moray eel is a fearsome sight. The discovery that morays can thrust a second pair of jaws out from their throat to wolf down prey whole increases their predatory reputation still further.

Many animals swallow their prey whole. Snakes come to mind, of course; but amphibians, lizards, birds and thousands of fish species can also attack, and gulp down, prey nearly as large as their head — imagine swallowing a peanut-butter sandwich or a salmon whole and you get some idea of how remarkable a feat that is. The methods by which animals do this vary from the suction feeding of fishes¹ to the 'unhinged-jaw' mechanism of snakes² and the inertial feeding of lizards and birds³. (For inertial feeding, picture a pelican thrusting its head upward to use gravity to choke down a large fish.)

These gulping mechanisms, along with most other vertebrate feeding habits that involve killing, dismembering and/or swallowing other animals, have generally been thoroughly investigated. On page 79 of this issue, however, Mehta and Wainwright⁴ document yet another tactic. They show how moray eels — elongated snake-like fishes that inhabit coral reefs and rocky intertidal habitats worldwide — drag a large item of prey into their gullet by using a second set of grasping jaws that they thrust forward from deep in their throat.

Accessory jaws positioned in the throat

are known as pharyngeal jaws, and are quite common among fishes. In many species, some of the bones that support the gills, called the branchial arches, have been modified into feeding tools that can filter prey from the water, crush and grind hard food such as snails or clams, and even grasp and tear softer prey before it is swallowed⁵.

Perhaps the most widely known pharyngeal jaws are found in freshwater cichlids and marine wrasses and parrotfishes. These fish families possess hard, toothy pharyngeal plates that are thought to have allowed a wide range of feeding habits to develop and promoted their evolutionary diversification⁶. But Mehta and Wainwright reveal⁴ an additional class of pharyngeal-jaw mechanism. They aptly term this the 'raptorial pharyngeal jaw', for its ability to reach far forward from its resting position in the pharynx, and grab the prey to transport it back towards the stomach.

The discovery of this mechanism in the reticulated moray eel (*Muraena retifera*) is notable in several respects. First, it is a classic example of discovery-based science, stemming from an inspirational "oh wow!" moment. Such moments

are crucial to the study of living organisms, for they complement the approach of testing *a priori* hypotheses with statistical analyses of large data sets. In this case, Mehta and Wainwright combined intellectual curiosity and visualization technology to reveal the moray eel's unusual behaviour. They had previously found⁷ that several types of eel do not use suction at all during feeding. This led them to search for alternative ways in which these predators could transport their prey into the oesophagus. By recording high-speed videos of eel feeding events in the laboratory, the mechanism became clear: the videos show the pharyngeal jaws projecting far forward into the mouth cavity to latch onto the food. Stills from one of the videos are shown as Figure 1, overleaf.

A second notable aspect of this research is the way the authors backed up their primary video records with a detailed anatomical study of the pharyngeal bones and muscles. They also performed X-ray fluoroscopic analysis (the same technique used in a 'barium swallow' to obtain pictures of the human gut) to examine the range of movement of the pharyngeal jaws. These data sets allowed a detailed and intriguing look at a novel feeding mechanism in this diverse and important group of animals.

As with all fascinating discoveries in animal behaviour and function, unexplained facets of the story remain. First, Mehta and Wainwright generalize their results obtained in a single species more broadly among eels. Although this may be valid for certain groups of morays within the 200-odd species in the moray family Muraenidae alone (and another 400 species in other eel groups), a more extensive comparative

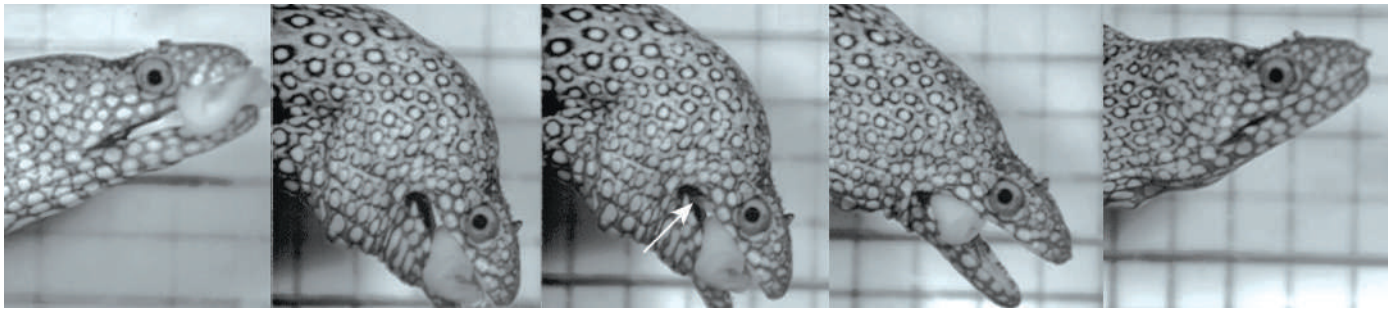


Figure 1 | Video action. These stills of a feeding moray eel come from the supplementary video evidence¹ that complements Mehta and Wainwright's detailed anatomical and X-ray fluoroscopic analyses. The pharyngeal jaws are seen only indistinctly here, but in the third frame (arrow) they shoot forward from within the pharynx to grasp the food item, which is then swiftly swallowed whole.

and evolutionary analysis of alternative strategies for prey capture and processing is needed to explain the origin and specialization of the raptorial pharyngeal-jaw mechanism.

The authors also compare the moray-eel transport mechanism to that of snakes. This is fair enough, as both groups are elongate, sinuous predators that can swallow their prey whole. But the ability of snakes to unhinge their jaws and alternately 'walk' the left and right jaws over a dead rodent to swallow it⁸ bears only a passing resemblance to the moray's ability to pull the prey down the gullet using a second set of jaws in the throat. These are different answers to a similar functional challenge, although further exploration of possible anatomical or

motor-pattern convergences in the two distantly related groups would certainly be worth while.

The prey-transport mechanism uncovered by Mehta and Wainwright⁴ will enhance the moray eel's fearsome reputation. But it also provides biologists with an important instance of a novel function in the skull of fishes that can be used to study the evolution of complexity. The spectacular assemblages of fishes on coral reefs, including eels, are among the greatest treasures of biodiversity. Now, when we snorkel or dive on a reef and see the fierce, toothy visage of a moray eel thrusting out of a crevice, we know that their raptorial pharyngeal jaws may be a primary reason for their success and diversity. ■

Mark W. Westneat is in the Department of

Zoology and Biodiversity Synthesis Center, Field Museum of Natural History, 1400 South Lakeshore Drive, Chicago, Illinois 60605-2496, USA.

e-mail: mwestneat@fieldmuseum.org

1. Westneat, M. W. in *Fish Biomechanics* (eds Shadwick, R. & Lauder, G. V.) 29–76 (Academic, New York, 2006).
2. Kley, N. J. *Am. Zool.* **41**, 1321–1337 (2001).
3. Smith, K. K. *J. Morphol.* **187**, 261–287 (1986).
4. Mehta, R. S. & Wainwright, P. C. *Nature* **449**, 79–82 (2007).
5. Wainwright, P. C. in *Fish Biomechanics* (eds Shadwick, R. & Lauder, G. V.) 77–101 (Academic, New York, 2006).
6. Mabuchi, K., Miya, M., Azuma, Y. & Nishida, M. *BMC Evol. Biol.* **7**, 10 (2007).
7. Mehta, R. S. & Wainwright, P. C. *J. Exp. Biol.* **210**, 495–504 (2007).
8. Cundall, D. & Greene, H. W. in *Feeding: Form, Function, and Evolution in Tetrapod Vertebrates* (ed. Schwenk, K.) 293–333 (Academic, San Diego, 2000).

PHYSICAL OCEANOGRAPHY

Super spin in the southern seas

Dean Roemmich

The southern oceans are generally considered as isolated systems, much like their northern counterparts. But a combination of historical data and new density profiles suggests that they may be connected on a global scale.

All Earth's major oceans contain a subtropical gyre, a vast circulation system spanning the entire ocean basin at mid-latitudes. These gyres have the crucial role in the climate system of exporting excess tropical heat polewards¹. Writing in *Geophysical Research Letters*, Ridgway and Dunn² analyse new and historical data from the Southern Hemisphere oceans and confirm a previous idea^{3,4} of an intermediate-depth connection linking the subtropical gyre circulations of the southern oceans. This opposes the conventional picture⁵ that the Southern Hemisphere mid-latitude circulation consists of distinct separated gyres within the Indian, Pacific and Atlantic oceans.

Ocean gyres are driven by the combination of low-latitude easterly trade winds and high-latitude westerlies, and so rotate clockwise in the Northern Hemisphere and anticlockwise in the Southern. The western boundary currents of these gyres, such as the Gulf Stream

in the North Atlantic, are among the strongest ocean currents in the world. Despite having the same basic form, there are pronounced differences between the northern and southern subtropical gyres. The North Atlantic and Pacific oceans are completely separated by the continents on their eastern and western sides, whereas the Southern Hemisphere is truly the 'ocean hemisphere'. The surface area between the Equator and 60° S is 84% ocean, and land masses form only partial barriers to the ocean currents.

The geography of the southern oceans made connections between them seem entirely plausible. Nevertheless, the confirmation of the existence of a 'supergyre' (an outer gyre surrounding the individual ocean gyres) connecting the Southern Hemisphere circulations represents a new understanding of the southern oceans. The centres of the South Pacific and South Indian gyres are located at 35° S–45° S at a depth of 1,000 m (Fig. 1a), with

the South Indian gyre extending far eastward south of Australia. New Zealand is effectively the western boundary of the South Pacific gyre between 35° and 45° S; without the islands there would be much more extensive connectivity between the Southern Pacific and Indian oceans. Despite this barrier, the relatively small amount of land mass in the Southern Hemisphere allows global winds to drive Southern Ocean circulation and hydrographic variability^{6,7} with minimal interruption.

The presence of the supergyre creates pathways for ocean-to-ocean mixing and exchange of water properties. The near-surface portion of the supergyre controls a significant amount of the planetary heat balance. For ocean surface waters, the westward limb of the South Pacific gyre (the South Equatorial Current) bifurcates at the coast of Australia, where part of the current turns northward across the Equator and feeds the Indonesian Throughflow (the flow from the Pacific to the Indian Ocean through the Indonesian archipelago). This large transport of warm water into the Indian Ocean, balanced by much colder eastward flow south of Australia, represents an immense heat exchange of about 4×10^{14} joules per second from the Pacific^{8,9}.

At intermediate depth, the existence of the southern supergyre has been suggested by simple models of wind-driven ocean circulation⁷. An agreement between models and data increases confidence in modelled climate trends in the Southern Hemisphere. These

D. ROEMMICH

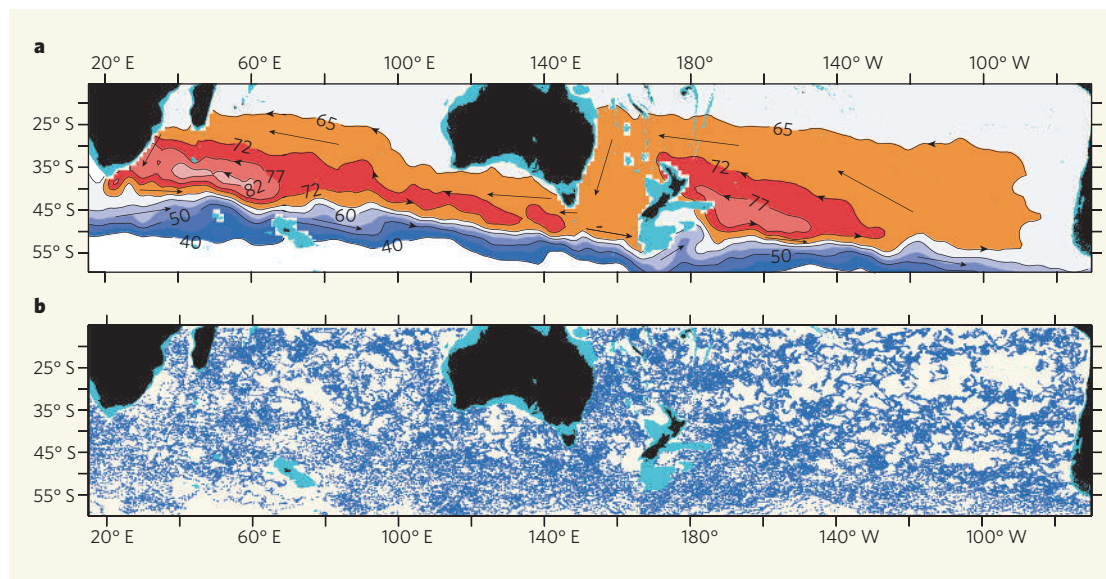


Figure 1 | Southern supergyre. **a**, The black arrows show ocean circulation and heat exchange between the southern Pacific and Indian oceans at a depth of 1,000 m. The orange band represents the area of supergyre circulation, and the pink and red bands show the distinct inner gyres. These gyres operate north of the Antarctic Circumpolar Current (blue band). The contour lines are dynamic height, which is generally derived from temperature and salinity profiles. **b**, The dark-blue dots represent the locations of Argo profiles used to generate **a**.

include a strengthening of the supergyre caused by increases in the subpolar westerly wind, a result of greenhouse warming and/or ozone depletion^{7,10}. Strengthened gyre and supergyre circulations may result in southward displacement of species' boundaries owing to warming of the Tasman Sea⁷ south of Australia, or a widespread decrease in primary productivity caused by a downward displacement of nutrient-rich, near-surface layers⁶.

Beyond illustrating a new model of southern circulation, the work of Ridgway and Dunn² highlights a revolution in ocean-observing capability over the past decade that is providing unprecedented views of the mean and time-varying ocean circulation. Inter-ocean connections are relatively subtle features, less obvious in the eddy-rich oceans than are the stronger circulations of individual ocean gyres. Observing such features, and their variability, requires data sets of geographical scope and quantity that did not exist until now. A search in the World Ocean Database¹¹ shows that from 1950 to 2000, a total of 22,809 temperature and salinity profiles to depths of at least 1,000 m were collected in the region in Fig. 1 (15° E to 70° W, 60° S to 15° S). The limited number of profiles and the temporal and geographical unevenness of those profiles make the historical data set grossly inadequate for detailed and accurate mapping of large-scale Southern Ocean circulation.

Since January 2004, however, the Argo global profiling float project¹² has collected more than 66,000 temperature and salinity profiles in the same region (Fig. 1b), including over 26,000 in the past 12 months. The deployment of the nearly 3,000 free-floating Argo instruments across the globe allows an unprecedented level of year-round data collection throughout the open ocean. Over the coming decade, our view of global ocean circulation and its variability will improve enormously as the new data set brings previously unknown features into sharp focus.

By combining the World Ocean Database and the new Argo profiles, Ridgway and Dunn²

not only confirm that the southern gyres are connected, exchanging heat and water masses on a global scale, but also show how much the new observational capabilities contribute to our understanding of the role of oceans in the climate system. The Southern Ocean, a critical region for climate research and one of the most hostile and inaccessible regions on Earth, now can be seen and studied by anyone with an Internet connection.

Dean Roemmich is at the Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0230, USA. e-mail: droemmich@ucsd.edu

1. Trenberth, K. & Caron, J. J. *Clim.* **14**, 3433–3443 (2001).
2. Ridgway, K. & Dunn, J. *Geophys. Res. Lett.* **34**, doi:10.1029/2007GL030392 (2007).
3. Reid, J. *Prog. Oceanogr.* **16**, 1–61 (1986).
4. Davis, R. E. J. *Phys. Oceanogr.* **35**, 683–707 (2005).
5. Tomczak, M. & Godfrey, J. S. *Regional Oceanography: An Introduction* (Pergamon, Tarrytown, NY, 1994).
6. Roemmich, D. et al. *J. Phys. Oceanogr.* **37**, 162–173 (2007).
7. Cai, W., Shi, G., Cowan, T., Bi, D. & Ribbe, J. *Geophys. Res. Lett.* **32**, doi:10.1029/2005GL024701 (2005).
8. Godfrey, J. S. *J. Geophys. Res.* **101**, 12217–12237 (1996).
9. Vranes, K., Gordon, A. & Field, A. *Deep-Sea Res. Pt II* **49**, 1391–1410 (2002).
10. Cai, W. *Geophys. Res. Lett.* **33**, doi:10.1029/2005GL024911 (2006).
11. www.nodc.noaa.gov/OC5/indprod.html
12. www-argo.ucsd.edu

CELL BIOLOGY

Taxi service for lipids

Anthony H. Futerman

With its role in intracellular protein transport already known, the FAPP2 protein has now also been implicated in lipid transfer and synthesis. What is more, these two FAPP2-mediated events seem to be linked.

The cellular pathways mediating the secretion of proteins such as hormones and antibodies were first described by George Palade and colleagues¹ in the 1960s. Many gaps have been filled since the heyday of their work, but the basic principles remain as originally described. Proteins that travel along the secretory pathway begin their life in the endoplasmic reticulum. They then move to the Golgi complex, where they are processed further and sorted for transport to their final destinations. The vehicles for protein transport are membrane vesicles, which bud off from one compartment and fuse with the next. By contrast, synthesis and transport of membrane lipids has received less attention. This is, in part, due to a widespread misperception that lipids are simply the glue that holds

proteins in place, and that their own transport within cells has little more than a passive role in secretion. On page 62 of this issue, D'Angelo *et al.*² challenge this view, showing that there is an unexpected link between lipid biosynthesis and protein secretion.

Lipids come in many shapes and sizes. The three most common types of lipid in mammalian cellular membranes are glycerolipids (of which phosphatidylinositol and its derivatives are the most famous), sterols and sphingolipids. Interest in sphingolipids has been renewed during the past two decades after the discovery of their function in molecular signalling pathways and in the formation of membrane microdomains known as lipid rafts. Complex sphingolipids are formed by the

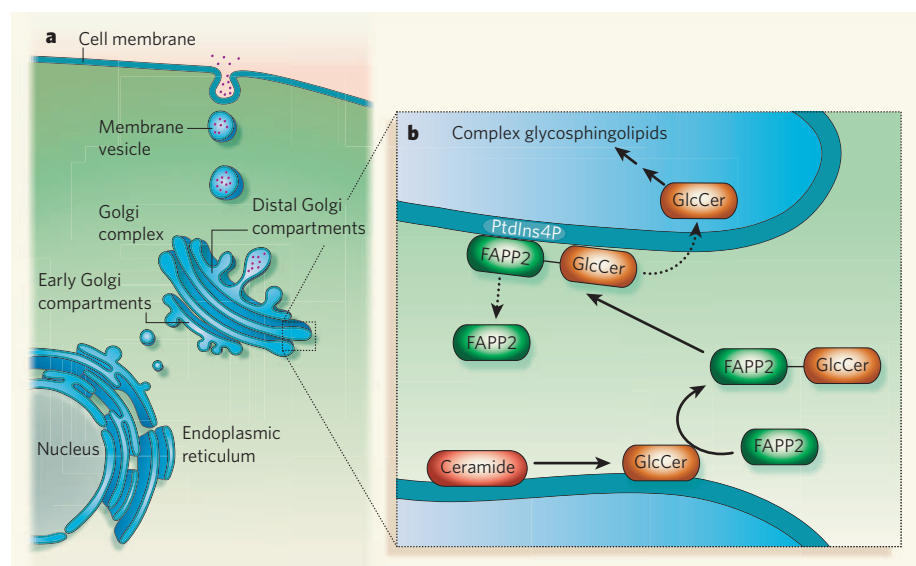


Figure 1 | Function of FAPP2 in glycosphingolipid biosynthesis. **a**, Secreted and membrane proteins, as well as most lipids, are synthesized in the endoplasmic reticulum, and are then transferred to the Golgi complex. While passing through the compartments of the Golgi complex — early and distal — they undergo further modifications and are sorted for transport to their final destinations. **b**, Glucosylceramide (GlcCer), which is synthesized from ceramide in the Golgi complex, has to be carried to more distal Golgi compartments for processing into complex glycosphingolipids. D'Angelo *et al.*² show that the FAPP2 protein mediates transport of GlcCer between Golgi compartments. It picks up GlcCer from the membrane of one Golgi compartment and, on reaching its destination, binds to the membrane by interacting with phosphatidylinositol 4-phosphate (PtdIns4P). Consequently, GlcCer is released and is translocated into the lumen of this compartment. Protein transport from distal Golgi compartments to the cell surface is spatially and temporally linked to ongoing GlcCer synthesis and transport, but the molecular details are not known.

addition of various head groups to a backbone ceramide molecule.

Glycosphingolipids are one such class of complex sphingolipid and are essential components of cell membranes, participating in various cellular processes including communication, differentiation, proliferation and adhesion. A bewildering number of these lipids can be formed by the sequential addition of sugar residues to sphingolipids. One of the simplest glycosphingolipids is glucosylceramide, which is formed by the addition of glucose to ceramide. Glucosylceramide is a precursor for the more complex glycosphingolipids.

Lipids, like membrane proteins, are synthesized mainly in the endoplasmic reticulum. Glycosphingolipids are an exception to this; although synthesis of the ceramide backbone occurs in the endoplasmic reticulum, complex glycosphingolipids are formed in the Golgi complex³. Glucosylceramide is even more unusual as, in contrast to more complex glycosphingolipids, which are all synthesized on the luminal (inside) surface of the Golgi complex, it is synthesized on the cytoplasmic side^{4,5}.

Sphingolipid aficionados had spent much time and energy attempting to rationalize why glucosylceramide is synthesized on the opposite side of the membrane from the more complex glycosphingolipids. D'Angelo *et al.*² might finally have found the answer. They show that glycosphingolipid synthesis depends on a cytoplasmic protein known as FAPP2, which was

previously shown^{6,7} to be associated with protein transport from the distal compartments of the Golgi complex — the compartments that are farthest away from the endoplasmic reticulum — to the cell membrane (Fig. 1a).

Previous work had shown that FAPP2 contains a domain that is similar to the glycosphingolipid-transfer protein⁸. So D'Angelo and colleagues asked whether FAPP2 could also transfer glucosylceramide: the answer is yes, in a highly specific manner. The group's findings indicate that FAPP2 binds to the surface of the Golgi complex facing the cytoplasm, picks up a molecule of newly synthesized glucosylceramide and delivers it to another, probably distal, Golgi compartment. Here, FAPP2 binds to the membrane by interacting with phosphatidylinositol 4-phosphate, a membrane lipid involved in the regulation of Golgi function⁹ (Fig. 1b).

On their own, these findings might seem esoteric. However, the bigger picture emerged when D'Angelo and co-workers addressed the question of whether there is a link between the function of FAPP2 in glycosphingolipid synthesis and its previously described^{6,7} role in regulating cellular trafficking. They found that, as expected, reintroducing normal FAPP2 into FAPP2-depleted cells restores both glucosylceramide transfer and protein transport. However, when a version of FAPP2 that could not mediate glucosylceramide transfer was reintroduced into FAPP2-depleted cells, protein transport could not occur.

These results clearly show that glucosylceramide synthesis in early Golgi compartments, as well as its transport by FAPP2 to distal Golgi compartments, is required for protein transport out of the distal compartments. The molecular details of how these pathways are connected have not yet been worked out, but presumably glucosylceramide, or a more complex glycosphingolipid, is required for an aspect of vesicle budding or formation at the distal Golgi compartment. Whether this is connected to the movement of glucosylceramide from the outside to the inside of the Golgi compartments, where it is metabolized further to complex glycosphingolipids, is also not known.

There might be further twists to this story, with the path of lipid transfer being perhaps less straightforward than the authors suggest. A recent study¹⁰ shows that FAPP2 can also mediate the backward transport of glucosylceramide from the Golgi complex to the endoplasmic reticulum, where it is processed to complex glycosphingolipids. But, although many details remain unresolved, together the two studies^{2,10} imply an unexpected function for FAPP2 in glycosphingolipid synthesis, and, as a consequence, in protein trafficking.

An exciting concept emerging from these studies is that events in the lumen of the Golgi compartments can be controlled from the cytoplasm; that is, FAPP2 performs its lipid-transfer activity in the cytoplasm, eventually binding to the Golgi membrane by interacting with a phosphoinositide lipid. This is reminiscent of other intracellular transport and signalling pathways. For instance, ligand binding to extracellular receptors initiates signalling cascades. These also often involve phosphoinositides, which regulate intracellular transport pathways. In his wish list¹, Palade stated, "there is a necessity to obtain comprehensive data on the chemistry and function of the different membranes of the secretory pathway and their interactions". The discovery^{2,10} of a set of extraordinarily complex feedback and feedforward loops to regulate glycosphingolipid synthesis, and their link to protein transport and secretion, is an exciting step towards achieving this goal.

Anthony H. Futerman is in the Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel.
e-mail: tony.futerman@weizmann.ac.il

1. Palade, G. *Science* **189**, 347–358 (1975).
2. D'Angelo, G. *et al.* *Nature* **449**, 62–67 (2007).
3. Futerman, A. H. & Riezman, H. *Trends Cell Biol.* **15**, 312–318 (2005).
4. Coste, H., Martel, M. B. & Got, R. *Biochim. Biophys. Acta* **858**, 6–12 (1986).
5. Futerman, A. H. & Pagano, R. E. *Biochem. J.* **280**, 295–302 (1991).
6. Godi, A. *et al.* *Nature Cell Biol.* **6**, 393–404 (2004).
7. Vieira, O. V. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 18556–18561 (2006).
8. Malinina, L., Malakhova, M. L., Teplov, A., Brown, R. E. & Patel, D. J. *Nature* **430**, 1048–1053 (2004).
9. Choudhury, R. R., Hyvola, N. & Lowe, M. *Biochem. Soc. Symp.* **72**, 31–38 (2005).
10. Halter, D. *et al.* *J. Cell Biol.* (in the press).

Q&A

D. HALLINAN/ALAMY; J. MONE/AP



M. LONGHURST/ALAMY; MEDICAL-ON-LINE/ALAMY

TECHNOLOGY

Biometric recognition

Anil K. Jain

Whether in passports, credit cards, laptops or mobile phones, automated methods of identifying people through their anatomical features or behavioural traits are an increasing feature of modern life.

Are biometric techniques the future of personal identification?

Yes, because unlike conventional recognition techniques such as passwords or ID cards, which are based on 'what you know' or 'what you have', biometric recognition is based on 'who you are': anatomical features such as face, fingerprint or iris, or behavioural traits such as signature or gait. This makes biometric technologies much more difficult to abuse than traditional methods of identification. Unlike passwords or ID cards, it is extremely difficult to guess, share, misplace, copy or forge biometric identifiers.

Why is the future arriving just now?

It's a combination of increased demand and increased supply. On the demand side, there are growing concerns about security threats and fraud. Governments want to keep track of who is entering and leaving their borders and receiving welfare payments; companies want to control who can enter their facilities, websites and proprietary databases. Crucially, public acceptance of the technology is also growing: citizens worried about identity theft are willing to use biometric systems for accessing laptops and mobile phones, and for making payments using credit cards at point-of-sale terminals. On the supply side, the explosion in the use of biometric techniques has been fuelled by the recent advent of compact and cheap sensors, and systems capable of fully automatic and 'real-time' identification (typically within a second).

How do these new technologies differ from familiar methods of biometric recognition such as fingerprinting?

Fingerprints have been used in forensics for about 100 years to identify repeat offenders and to establish the identity of a criminal from prints left at the scene of a crime. But in traditional fingerprint identification, a human expert is generally in the loop (so to speak), to make the final determination of identity from a candidate list generated by an automatic system. Other methods of identification used in forensics — analysis of DNA, hair and fibre samples, for example — are not fully automated either and take a long time (hours to days) to make the identification. That's OK for criminal investigations, but no good for many commercial applications.

How are fingerprinting technologies being updated?

The traditional method of capturing fingerprints, still practised by police and government agencies, is based on 'inked impression on paper', called rolled fingerprints. But electronic sensors based on optical, solid-state, thermal, ultrasound and multispectral technologies are now available. These sensors generate digital images from differences in the physical properties (such as reflectance and capacitance) of ridges and valleys as the finger surface touches a plate (Fig. 1, page 40). 'Touchless' variants record an image of the finger surface directly using one or more digital cameras.

Many of these sensors are extremely compact and cheap, enabling them to be embedded in consumer electronic products such as mobile phones, personal digital assistants and laptops.

Iris recognition is a buzz phrase at the moment. What does this involve?

Iris recognition is based on the analysis of unique and stable texture patterns that are visible within the iris (the coloured portion) of the human eye. Video-camera technology is used to record an image of the eye, and the iris region is localized after digitally removing the pupil, eyelids and eyelashes. The phase information of the digitally filtered iris image (which represents the unique wavy pattern of each iris), is computed and stored as a string of bits. Recognition is performed by comparing two such bit sequences.

What are the elements of a biometric recognition system?

They all have four features in common (Box 1). First, there is the sensor, to capture or read the biometric trait — a fingerprint, iris, signature, voice trait or similar. Then there's the feature extractor, to extract some salient characteristics of the trait for recognition; the enrolment database, where the biometric features (also called templates) of all the enrolled users of the system are stored; and the matcher, which compares an input biometric sample with the templates in the database.

How is the identification performed?

We can distinguish between two cases: positive and negative recognition. Positive recognition is easy. Say you want to log in to your laptop with a built-in fingerprint reader. Instead of entering your password, you place your finger on the sensor. If the features extracted from your fingerprint match the fingerprint template associated with your ID in the laptop's enrolment database, access is granted.

And how does negative recognition work?

Negative recognition is more challenging. Suppose you want a new driver's licence. Typically, you have to produce one or more forms of paper identification such as a birth certificate or a passport. But these can be relatively easily forged. How does the licensing authority know that you do not already have a licence under a different name? A biometric database allows the licence issuer to find this out by matching a biometric trait (your face, for example) with that of all the individuals in the database who have already been issued with a licence. There are many such examples in which 'multiple enrolments' by the same person need to be detected — issuing passports and disbursing welfare payments, for example.

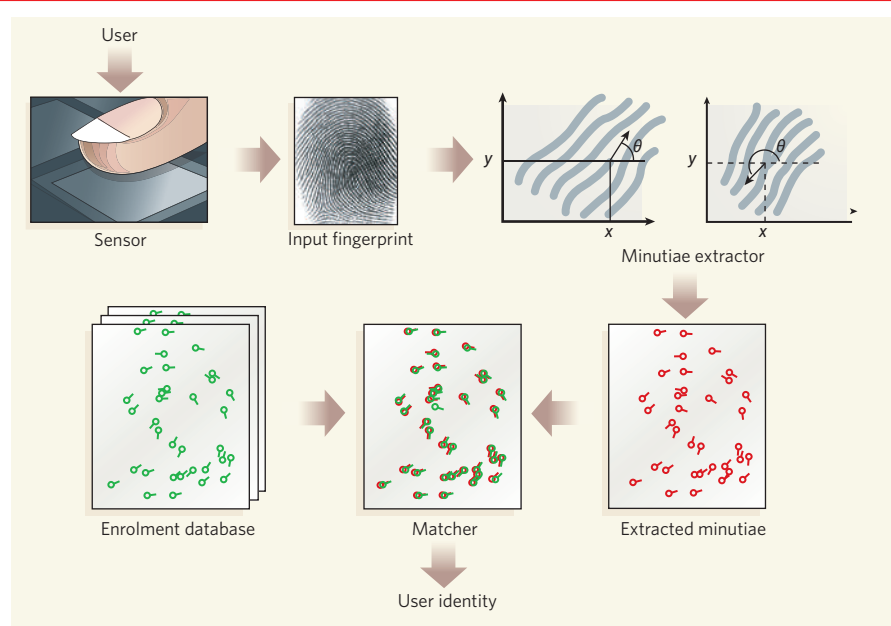
Have biometric systems been successfully deployed on a large scale?

Certainly. An example of a high-throughput positive-recognition system is the multispectral fingerprint sensor used by the Walt Disney World Resort in Orlando, Florida, to prevent ticket fraud. Every visitor to the resort must provide his or her index finger at the turnstile along with the ticket. That fingerprint gets linked to that particular ticket, so if you visit the resort again (either later the same day or on a different day if you have a multiple-entry ticket) you must present the same finger that was used to validate the ticket. The Disney system can handle a large number of visitors (around 100,000 every day) efficiently and — an important point, this — it works equally well in all weathers. Another example is the UK government's Iris Recognition Immigration System (IRIS), a positive-recognition system based on iris scanning, which allows enrolled travellers to bypass normal immigration channels at major airports.

Are there working negative-recognition systems?

The most high-profile example of a negative-recognition system is perhaps the United States Visitor and Immigrant Status Indicator Technology (US-VISIT) system used by the Department of Homeland Security. Every visitor to the United States must now provide fingerprint images of left and right index fingers and a facial image at their port of entry. The two fingerprints are matched against a watch-list of 2.5 million records in less than 10 seconds. More than 75 million visitors have

Box 1 | Automated biometric identification



A typical biometric recognition system, here an automated fingerprint identification system, has two stages of operation: the enrolment stage and the identification stage. In the enrolment stage, a sensor captures a digital image of a legitimate user's fingerprint. Its salient features, or minutiae, are extracted and stored as a template in an enrolment database.

These minutiae take the form of locations (x and y coordinates) and orientations (θ) of abrupt ends and junctions of fingerprint ridges. In a high-quality image, there are typically 20–70 such minutiae, depending on the size of the sensor surface and how users place their finger on the sensor.

During identification, fingerprint minutiae are extracted from a query print in the same way and compared with the minutiae of the templates stored in the enrolment database. Variations in placement and pressure mean that template and query fingerprints must be aligned before matching. The number of minutiae that have similar x , y and θ coordinates forms a basis for determining the identity of the user.

In iris recognition, phase information from the pattern of the iris takes the place of minutiae and is stored in a database in the form of a barcode. For face recognition, the distribution of pixel intensities and positions of features in a facial image are extracted for comparison with stored templates. **A.K.J.**

been processed through this system since its inception in January 2004, and about 1,000 have been denied entry.

Where has biometric technology been tried and found wanting?

Surveillance at public places such as airports and busy streets is a perennial problem for biometric recognition. Although surveillance cameras can be used to detect suspicious behaviour or events, capturing facial images and thereby identifying known criminals and hooligans is more difficult. Face-matching algorithms generally use statistical techniques to analyse the distribution of pixel intensities in a face image and measure the relative positions of different features (eyes, nose and so on). But state-of-the-art face-recognition algorithms are not very accurate when the light is not good, when only a partial face image is available, or when a person's appearance has changed since the entry in the database, through, for example, ageing, a different expression or the addition of glasses or a beard.

How is the accuracy of a biometric recognition system measured?

Through two figures: the false reject rate (FRR),

the frequency with which a genuine user is not correctly recognized and hence denied access; and the false accept rate (FAR), the frequency with which an impostor is accepted as a genuine user. Of course, these two metrics are not independent for a given system. A system that accepts all the right people, and so has a desirably low FRR, might not reject all the wrong people, and so could have a higher FAR.

How accurate are current systems?

Current fingerprint-recognition systems can provide an FRR of up to 0.01% (1 in 10,000) at an FAR of 0.1% (1 in 1,000). Of course, the actual performance of a biometric system depends on several factors, including the specific biometric trait, characteristics of the sensor used to capture it, the number and characteristics of the people enrolled in the database, as well as various environmental factors (indoors or outdoors, temperature, humidity, and so on). And where the emphasis on accuracy lies depends on the specific application: Disney World, for example, will want as low an FRR as possible, so as not to upset customers unnecessarily, at the expense of a higher FAR. For the US-VISIT system, the opposite is true: a low FAR is needed (to keep potential

criminals and terrorists out) at the expense of a higher FRR.

What can be done to decrease errors?

Efforts are afoot to design better biometric sensors/readers, to improve algorithms to extract features from raw biometric data and to match two biometric samples quickly and accurately. Another research track being pursued specifically to decrease recognition error is to fuse information from several independent biometric sources. The US-VISIT scheme, with its two-fingerprint strategy, is an operational example of such a 'multibiometric' system. In the future, the fingerprint images could be supplemented with data from face-recognition software. Several different types of multibiometric systems can be envisaged (Fig. 2).

Are some biometric traits better than others?

Yes. Fingerprint and iris have the smallest error rates. But error rate is only one consideration. Other factors include system cost, size of the scanner, integration into existing security infrastructure, and user comfort and perception. So there's no perfect biometric trait.

Are people's fears about biometric technology justified?

Concerns can be classified into three broad categories. First, there is security: that a biometric system can itself be attacked or compromised (for example, by providing a fake input fingerprint). Second, there is the issue of privacy: that governments and states might use the technologies to track and snoop on people, or that a person's biometric template might be stolen and abused. Third, there are cultural and religious objections. Concerns relating to security and privacy issues are justified, and are the subject of research. As far as privacy and cultural objections are concerned, government regulation and public education will be required if full acceptance is to be achieved.

Can biometric security realistically be breached?

One way to attack a biometric system particularly beloved of film-makers and authors is to employ a spoof biometric trait (an artificial or dead finger, or a face mask, for example). This is a serious concern, but new fingerprint

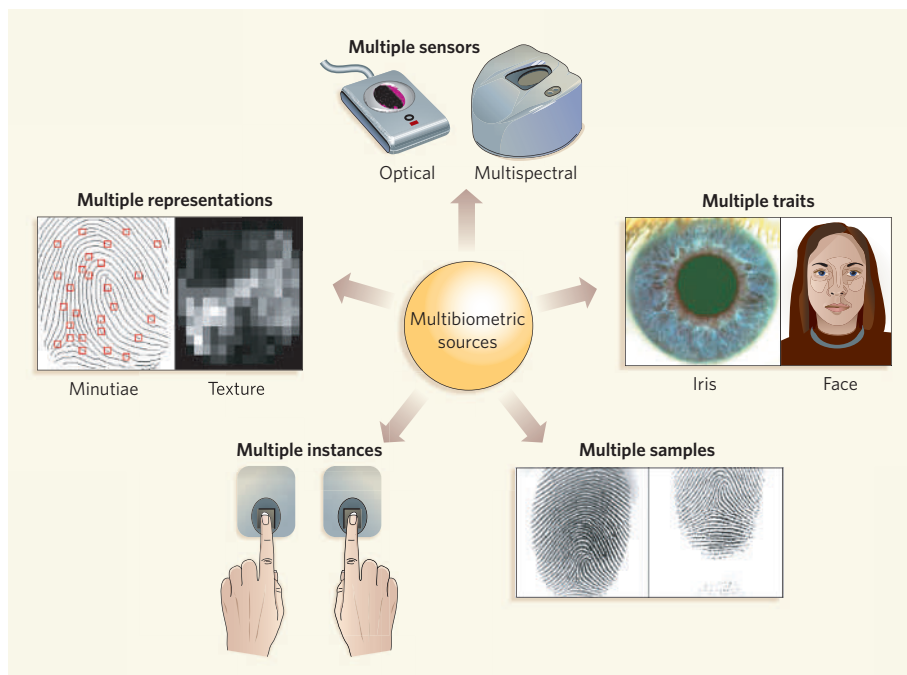


Figure 2 | Multibiometric identification. Sources of information in a multibiometric system include (clockwise from the top) using different sensors to capture the same biometric trait; using more than one independent trait; taking more than one sample of the same trait (for example, the same fingerprint or iris scan); taking multiple instances of a trait (prints of both the left and right index finger, or the irises of right and left eyes); and more than one type of representation of the same trait. The US-VISIT immigration scheme is a multibiometric system based on multiple instances (two fingerprints) with the future potential also to take a multiple-trait approach, fusing fingerprint data with face-recognition technology.

sensors can detect whether the finger placed at the sensor is living or not by using, for example, a multispectral imaging technique to measure how much light the finger absorbs, or by measuring the finger's electrical conduction properties using electric-field sensors.

What about hacking into the underlying computer systems?

The most sensitive parts of a biometric system are the enrolment database, an attack on which constitutes both a security and a privacy threat, and the channels used to transmit information between the different elements of the system. Both can be protected through clever combinations of biometrics with cryptographic techniques to prevent hackers intercepting, relaying or modifying information. Alternatively, to eliminate the danger of interception in transmission, the entire biometric system (sensor, feature extractor, database and

matcher) can be built on a single chip or a 'smart card', so that no biometric data ever leave the card or the chip.

Will concerns about its security stop the onward march of biometric recognition?

No, because whatever its problems, biometric recognition offers greater security and convenience than traditional methods of person recognition based on official documents, PINs and passwords. In some applications, such as access to computer systems, biometrics can replace or supplement existing methods to improve recognition accuracy. In others, such as issuing passports and driving licences, biometrics is the only viable approach for determining whether an individual has already been issued these documents under a different name. Like it or loathe it, the technology is here to stay. ■

Anil K. Jain is in the Department of Computer Science and Engineering, Michigan State University, East Lansing, Michigan 48824-1226, USA.
e-mail: jain@cse.msu.edu

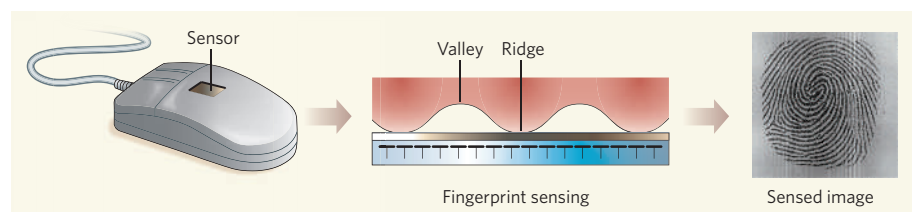


Figure 1 | A capacitive fingerprint sensor. When a user places a finger against a silicon chip containing an array of microcapacitor plates, a small electric charge arises in the insulating air gap, the magnitude of which depends on the distance between the finger and the plates. The capacitance values of different plates, converted into pixel intensities, form a digital image of the ridges and valleys of the fingerprint. Such fingerprint-sensing chips cost only about US\$5, and are compact enough to be embedded in mobile phones, key fobs and hand-held computers.

FURTHER READING

Jain, A. K., Bolle, R. & Pankanti, S. (eds) *Biometrics: Personal Identification in Networked Society* (Springer, Heidelberg, 2006).
Maltoni, D. et al. *Handbook of Fingerprint Recognition* (Springer, Heidelberg, 2003).
Daugman, J. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1148-1161 (1993).
Ross, A., Nandakumar, K. & Jain, A. K. *Handbook of Multibiometrics* (Springer, Heidelberg, 2006).
Rowe, R. K. et al. *Proc. SPIE* **5694**, 90-99 (2005).

Dragging of inertial frames

Ignazio Ciufolini¹

The origin of inertia has intrigued scientists and philosophers for centuries. Inertial frames of reference permeate our daily life. The inertial and centrifugal forces, such as the pull and push that we feel when our vehicle accelerates, brakes and turns, arise because of changes in velocity relative to uniformly moving inertial frames. A classical interpretation ascribed these forces to acceleration relative to some absolute frame independent of the cosmological matter, whereas an opposite view related them to acceleration relative to all the masses and ‘fixed stars’ in the Universe. An echo and partial realization of the latter idea can be found in Einstein’s general theory of relativity, which predicts that a spinning mass will ‘drag’ inertial frames along with it. Here I review the recent measurements of frame dragging using satellites orbiting Earth.

Inertial frames are systems in which any isolated body, not subject to any interaction, is at rest or moving with constant velocity. However, what determines an inertial frame? And, in general, what is the origin of inertia? In the mechanics of Galileo Galilei and Isaac Newton (our everyday experience, in general) an inertial frame has an absolute existence, uninfluenced by the matter in the Universe. In Einstein’s general theory of relativity^{1–3}, the inertial and centrifugal forces are due to our accelerations and rotations with respect to the so-called local inertial frames which, in turn, are determined, influenced and dragged by the distribution and flow of mass-energy in the Universe. In particular, they are dragged by the motion and rotation of nearby matter; this general relativistic phenomenon is called frame-dragging (ref. 4, and see page 544 of ref. 1 for example) and represents in Einstein’s theory the remnant of the ideas of Ernst Mach on the origin of inertia. Mach thought that centrifugal and inertial forces are due to rotations and accelerations with respect to all the masses in the Universe; this is known as Mach’s principle^{5,6}.

During the past century the general theory of relativity gave rise to an experimental triumph^{3,7–9}. On the one hand, a number of key predictions of Einstein’s gravitational theory have been experimentally confirmed with impressive accuracy⁹ (Table 1). On the other hand, the general theory of relativity today has practical applications in space research³, geodesy, astronomy and navigation in the Solar System, from the global positioning satellite system to the techniques of very long baseline interferometry (VLBI) and satellite laser ranging, and is a fundamental component for the understanding of some astrophysical and cosmological observations.

Frame-dragging has an intriguing influence on the flow of time around a spinning body (Fig. 1). Indeed, synchronization of clocks all around a closed path near a spinning body is not possible^{10,11} in any rigid frame not rotating relative to the ‘fixed stars’, because light co-rotating around a spinning body would take less time to return to a starting point (fixed relative to the ‘distant stars’) than would light rotating in the opposite direction^{10–12}. Frame-dragging may be usefully described by a formal analogy of the general theory of relativity with electrodynamics^{3,13,14} (Fig. 2).

In summary, the general theory of relativity has so far passed all the tests it has been put to, including a confirmation of the existence of frame-dragging by spin with an accuracy of approximately 10% (refs 15 and 16). Among the biggest remaining challenges are the direct detection of gravitational waves, although we already know from the rate of change of the orbital period of the binary pulsar PSR B1913+16 that this system emits gravitational waves at the

predicted level⁹, and the merging of the general theory of relativity with quantum mechanics¹⁷, which may even solve the problem of the spacetime singularities^{18,19}, events predicted by Einstein’s gravitational theory in which the known physical theories cease to be valid and the spacetime curvature diverges. In the following I focus on the measurement of frame-dragging.

Table 1 | Main 2007 tests of the general theory of relativity

Phenomenon or principle tested	Method and 2007 experimental limit
Weak equivalence principle (test-particles fall with the same acceleration; this is at the foundations of geometrical (metric) theories of gravitation)	Laboratory experiments (accuracy of the order of 10^{-13}) Lunar laser ranging (accuracy of the order of 10^{-13})
Strong equivalence principle (this is at the foundations of the general theory of relativity)	Lunar laser ranging (accuracy of less than 10^{-3})
Gravitational time dilation or gravitational redshift (relative slowing down of clocks near a mass)	Gravity Probe A (with a clock on the ground and one on a rocket; accuracy of the order of 10^{-4})
Deflection of photons’ path and travel time-delay of electromagnetic waves, or Shapiro time-delay, by a mass	VLBI (accuracy of the order of 2×10^{-4}) Cassini spacecraft tracking (accuracy of the order of 10^{-5})
Perihelion advance of Mercury	Mercury radar ranging (accuracy of the order of 10^{-3})
Periastron advance, time dilation, time delay, rate of change of the orbital period (accurately explained by the loss of energy due to the emission of gravitational waves from a binary system) and other relativistic parameters (these effects are characterized by strong gravitational field inside a pulsar)	Binary pulsar PSR 1913+16 Other binary pulsars
Lense–Thirring effect, or frame-dragging of a gyroscope by the spin of a body	LAGEOS and LAGEOS2 laser ranging (accuracy of the order of 10^{-1}) Gravity Probe B (it might be detected by further GP-B data analysis)
Geodetic precession, or de Sitter effect (dragging of a gyroscope due to its motion in a static gravitational field)	Lunar laser ranging (accuracy of the order of 6×10^{-3}) Gravity Probe B (accuracy of the order of 1.5×10^{-2} ; it should be improved by further Gravity Probe B data analysis) Binary pulsars

For a detailed description see refs 7, 8 and 3; for an update see ref. 9.

¹University of Salento and INFN Sezione di Lecce, Via Monteroni, 78100 Lecce, Italy.

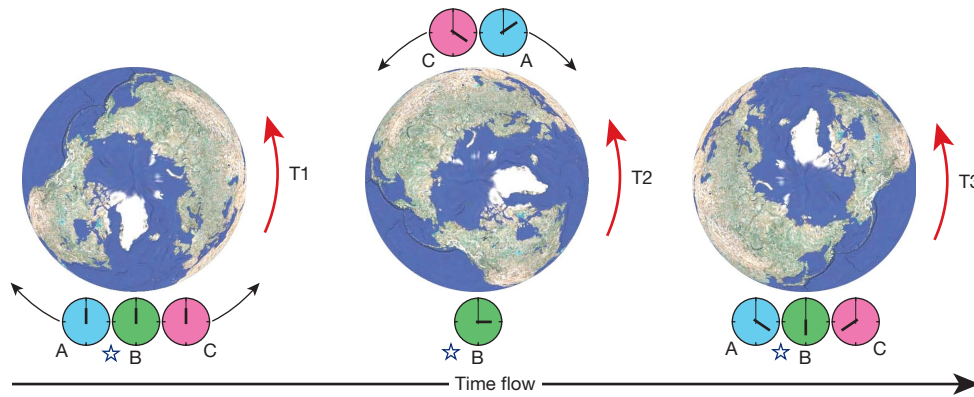


Figure 1 | Frame-dragging effects on clocks by a rotating mass. If two clocks, or twins A and C, fly all around a spinning body, even very slowly, and a third one B awaits them at the starting point, fixed relative to the 'distant stars' (a 'fixed star' is shown in blue, and T1, T2 and T3 are three consecutive instants of time), then when they meet again, the twin A that was travelling in the direction opposite to the rotation of the central body, would be younger relative to the twin B awaiting at the starting point. On the other hand twin C, travelling in the same direction of rotation of the body, would be older with respect to the standing twin B and to the twin A rotating in the opposite direction^{10–12}. For this time dilation, due to the spin of the central body, to

occur, they would not need to move near the speed of light (as in the case of the well-known 'twin-paradox' of special relativity). For example, if two such twins meet again, having flown arbitrarily slowly around the whole Earth in opposite directions on the equatorial plane and exactly at the same altitude, the difference in their ages due to the Earth's spin would be approximately 10^{-16} s (for an altitude of about 6,000 km), which would be in principle detectable if not for the other, much larger, relativistic clock effects. However, frame-dragging does produce relevant effects on light and matter around a rotating black hole⁸¹. (Earth's image by NASA and Google Earth.)

The LAGEOS satellites

A spinning gyroscope defines very accurately by its rotation an axis fixed relative to the local inertial frames. Similarly, the orbital plane of a planet, moon or satellite is a huge gyroscope that 'feels' general relativistic effects. One of the early triumphs^{3,7,8} of the general theory of relativity was the prediction of the precession of Mercury's perihelion by the mass of the Sun, whose deviation from the newtonian result was becoming something of an embarrassment to astronomers. In 1916, using the general theory of relativity, de Sitter²⁰ also calculated the additional frame-dragging effect consisting in the much smaller Mercury's perihelion shift due to the Sun's spin (Box 1).

In 1918, Lense and Thirring²¹ formulated the weak-field and slow-motion description of frame dragging on the orbit of a test particle around a spinning body, now known as the Lense–Thirring effect (Box 1 and Fig. 3). But frame-dragging is extremely small for Solar System objects, so to measure its effect on the orbit of a satellite

we need to measure the position of the satellite to extremely high accuracy.

The accurate measurement of distances is a fundamental task in science and technology. Laser-ranging is the most accurate technique for measuring distances to the Moon^{22,23} and to artificial satellites such as LAGEOS (laser geodynamics satellite)²⁴. Short-duration laser pulses are emitted from lasers on Earth and then reflected back to the emitting laser-ranging stations by retro-reflectors on the Moon or on artificial satellites. By measuring the total round-trip travel time we are today able to determine the instantaneous distance of a retro-reflector on the LAGEOS satellites with a precision of a few millimetres²⁵.

LAGEOS²⁴ was launched by NASA in 1976 and LAGEOS2 was launched by the Italian Space Agency and NASA in 1992, at altitudes of approximately 5,900 km and 5,800 km respectively. The LAGEOS satellites' orbits can be predicted, over a 15-day period, with an uncertainty of just a few centimetres^{16,25}. The Lense–Thirring drag

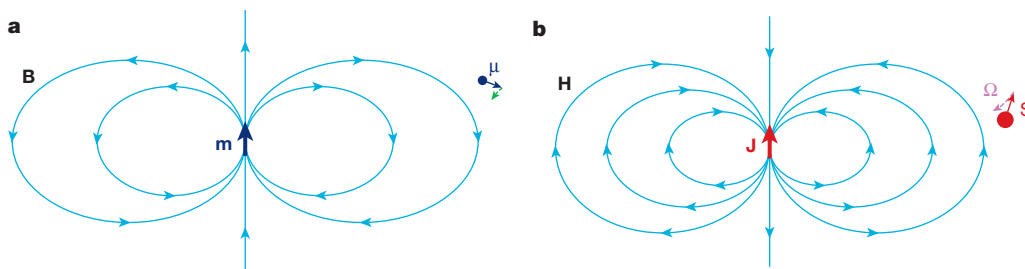


Figure 2 | Frame-dragging and the gravitomagnetic analogy of the general theory of relativity with electrodynamics. In the general theory of relativity, freely falling test-gyroscopes, that is, sufficiently small and accurate spinning tops, determine the axes of the local, non-rotating, inertial frames^{1–3}, where the equivalence principle holds—that is, where the gravitational field is locally 'unobservable' and all the laws of physics are the laws of the special theory of relativity². Therefore, if we rotate with respect to these gyroscopes, we feel centrifugal forces, even though we may not rotate at all with respect to the 'distant stars', contrary to our everyday intuition. Indeed, a gyroscope is dragged by spinning masses, that is, its orientation changes with respect to the 'distant stars'. Frame-dragging phenomena, which are due to mass currents and mass rotation, have been called gravitomagnetism^{3,13} because of a formal analogy of electrodynamics with the general theory of relativity (in the weak field and slow motion approximation). Whereas an electric charge generates an electric field and a current of electric charge generates a magnetic field, in newtonian gravitational theory the mass of a body generates a gravitational

field but a current of mass, for example the rotation of a body, would not generate any additional gravitational field. On the other hand, Einstein's gravitational theory predicts that a current of mass would generate a gravitomagnetic field that would exert a force on surrounding bodies and would change the spacetime structure by generating additional curvature⁸¹. The gravitomagnetic field generates frame-dragging of a gyroscope, in a similar way to the magnetic field producing the change of the orientation of a magnetic needle (magnetic dipole). Indeed, in the general theory of relativity, a current of mass in a loop (that is, a gyroscope) has a behaviour formally similar to that of a magnetic dipole in electrodynamics, which is made of an electric current in a loop (Box 1). In panel a I show the magnetic field **B** generated by a magnetic dipole **m** and a test magnetic dipole **μ**, that is, a magnetic needle, which tends to be aligned along **B**. In panel b is the gravitomagnetic field¹⁴ **H** generated by the spin **J** of a central body and frame-dragging **Ω** of a test gyroscope **S**.

Box 1 | Frame-dragging in weak gravitational field and slow motion

The precession, with rate Ω_{L-T} , of the longitude of the nodal line of a test-particle, that is, of its orbital angular momentum vector, is:

$$\Omega_{L-T} = \frac{2J}{a^3(1-e^2)^{3/2}}$$

where J is the angular momentum of the central body, a the semi-major axis of the orbiting test-particle and e its orbital eccentricity.

The rate of change ω_{L-T} of the longitude of the pericentre of a test-particle, that is, of the so-called Runge–Lenz vector, is:

$$\omega_{L-T} = \frac{2J}{a^3(1-e^2)^{3/2}}(\hat{J} - 3 \cos I \hat{I})$$

where I is the orbital inclination, that is, the angle between the orbital plane of the test-particle and the equatorial plane of the central object, \hat{J} is the angular momentum unit-vector of the central body and \hat{I} the orbital angular momentum unit-vector of the test-particle.

The precession Ω_S of the spin axis of a test-gyroscope by the angular momentum J of the central body is:

$$\Omega_S = \frac{3(J \cdot \hat{r})\hat{r} - J}{r^3}$$

where \hat{r} is the position unit-vector of the test-gyroscope and r is its radial distance from the central body.

The geodetic precession Ω_{geodetic} of a test-gyroscope due to its velocity v , orbiting at a radial distance r from a mass M , is:

$$\Omega_{\text{geodetic}} = -\frac{3M}{2r^2} \mathbf{v} \times \hat{r}$$

of the orbital planes of LAGEOS and LAGEOS2 (Box 1) is^{26,27} approximately 31 milliarcseconds per year, corresponding at the LAGEOS altitude to approximately 1.9 m yr^{-1} . Using laser-ranging we can determine their orbits with an accuracy of a few centimetres, so the Lense–Thirring effect can be measured very accurately on the LAGEOS satellites' orbits if all their orbital perturbations can be modelled well enough^{26–30,82}. On the other hand, the LAGEOS satellites are very heavy spherical satellites with small cross-sectional areas, so atmospheric particles and photons can only slightly perturb their orbits³¹ and especially they can hardly change the orientation of their orbital planes^{28–30,32,82}.

By far the main perturbation of their orbital planes is due to the Earth's deviations from spherical symmetry³³. In particular, the flattening of the Earth's gravitational potential produces a large perturbation of the LAGEOS node^{28–30,82}. But thanks to the observations of the geodetic satellites, the Earth's shape and its gravitational field are extremely well known. For example, the flattening of the Earth's gravitational potential is today measured³⁴ with an uncertainty of only about one part in 10^7 . To eliminate the orbital uncertainties due to the errors in the Earth's gravity models, the use of both LAGEOS and LAGEOS2 was proposed²⁷. However, it was not easy to confidently assess the accuracy of the earlier measurements^{35,36} of the Lense–Thirring effect with the LAGEOS satellites, given the limiting factor of the uncertainty of the gravity models available in 1998.

The problem^{37,38} of the uncertainties in the Earth's gravity field was overcome in March 2002 when the twin GRACE (gravity recovery and climate experiment)^{39,40} spacecraft of NASA were launched in a polar orbit at an altitude of approximately 400 km and about 200–250 km apart. The spacecraft range to each other using radar and they are tracked by the global positioning satellites. The GRACE satellites have greatly improved our knowledge of the Earth's gravitational field. Indeed, by using the two LAGEOS satellites and the GRACE Earth gravity models³⁴, the orbital uncertainties due to the modelling errors in the non-spherical Earth's gravitational field are only a few per cent of the Lense–Thirring effect¹⁶. In 2004, nearly eleven years of laser-ranging data were analysed. This analysis resulted in a measurement of the Lense–Thirring effect with an accuracy^{15,16} of approximately 10%; the

main error source was the uncertainty in some of the Earth's axially symmetric departures from sphericity (see Figs 3 and 4). After 2004, other accurate Earth gravity models have been published using longer GRACE observations. The LAGEOS analyses have been recently repeated with these models, over a longer period and by using different orbital programs independently developed by NASA Goddard and the GeoForschungsZentrum (GFZ) Potsdam. These recent frame-dragging measurements⁴¹, by a team from the universities of Salento, Rome, Maryland, NASA Goddard and the GFZ Potsdam, have improved the precision of the 2004 LAGEOS determination of the Lense–Thirring effect. No deviations from the predictions of the general theory of relativity have been observed. The laser-ranged satellite LARES (laser relativity satellite, Italian Space Agency), should in future provide an improved test of the Earth's gravitomagnetism with accuracy of the order of 1%.

Gravity Probe B

In 1959 and 1960, an experiment to test the general relativistic drag of a gyroscope was suggested^{42–44}. On 20 April 2004, after more than 40 years of preparation, the Gravity Probe B spacecraft was finally launched in a polar orbit at an altitude of about 642 km (Fig. 5). The Gravity Probe B mission^{45,46} (see an update of Gravity Probe B at <http://einstein.stanford.edu/>) consisted of an Earth satellite carrying four gyroscopes and one telescope, and was designed to measure the relativistic precessions of the four test-gyroscopes with respect to the distant 'fixed' stars. Whereas frame-dragging affects the orbital plane of the LAGEOS satellites, on Gravity Probe B it acted on its small gyroscopes. The two relativistic effects sought for are geodetic precession and frame-dragging by the Earth angular momentum.

The general theory of relativity predicts (Box 1) that the average frame-dragging precession of the four Gravity Probe B gyroscopes by the Earth's spin will be about 39 milliarcseconds per year, that is, 0.000011 degrees per year, about an axis contained in Gravity Probe B's polar orbital plane. The geodetic precession, due to the motion of the four gyroscopes around the Earth mass, is a much larger drift of approximately 6,600 milliarcseconds per year, that is, 0.0018 degrees per year, about an axis orthogonal to Gravity Probe B's orbital plane

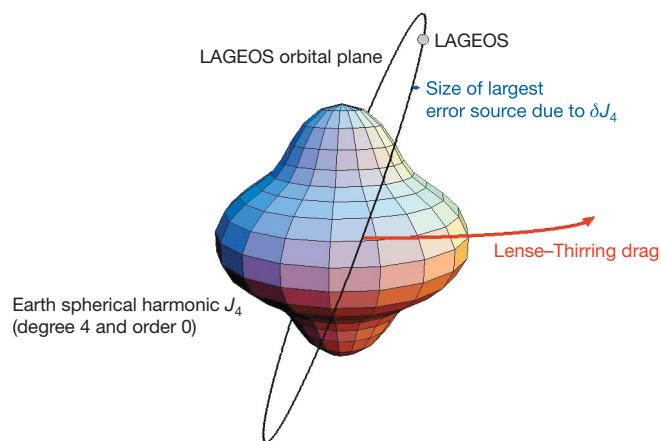


Figure 3 | The Lense–Thirring effect on the orbital plane of a test-particle. The Lense–Thirring precession of the orbital plane of a test-particle by the spin of a central body is represented by the big red arrow. Also shown is the Earth deviation from spherical symmetry (enhanced, and so not to scale) described by the so-called even zonal harmonic of degree four, J_4 . The uncertainty in its static part is the largest source of error in the present measurement of frame-dragging using the LAGEOS satellites. The maximum precession of the LAGEOS orbital plane due to the uncertainty in J_4 , that is, the nodal precession error due to δJ_4 , is represented by the blue arrow; this error and the Lense–Thirring effect are drawn to scale; indeed, the nodal uncertainty due to the error δJ_4 corresponds, according to the 2004 GRACE (GFZ) Earth gravity model³⁴, to only 3% of the Lense–Thirring effect.

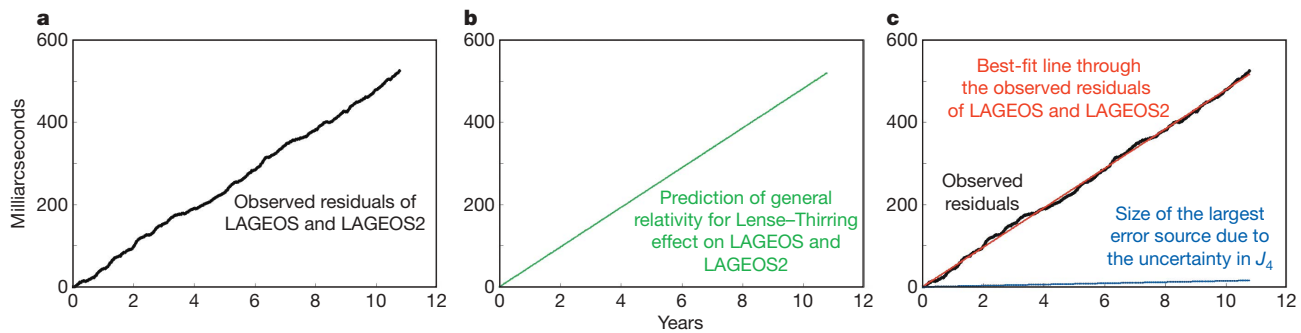


Figure 4 | The Lense–Thirring effect measured via the LAGEOS satellites^{15,16} in 2004 and its theoretical value predicted by the general theory of relativity. **a**, The black solid line represents the observed residuals of the combined nodal longitudes of the LAGEOS satellites. **b**, The green solid line is the theoretical, Lense–Thirring, prediction of general relativity. **c**, The red solid line is the best-fit line through the observed residuals (in black) and the blue solid line represents the uncertainty in the combined

nodal longitudes of the LAGEOS satellites from the largest error source due to the uncertainty in the Earth’s even zonal harmonic of degree four, J_4 (corresponding to approximately 3% of the Lense–Thirring effect according to the 2004 GRACE (GFZ) Earth gravity model³⁴; see Fig. 3). The observed slope of the red line is 0.99 ± 0.1 , where 1 is the prediction of the general theory of relativity (green line) and the ± 0.1 uncertainty is the estimated total systematic error.

(Box 1 and Fig. 5). For a polar satellite, geodetic precession and average Lense–Thirring effect are then orthogonal to each other. Gravity Probe B was designed to measure the geodetic precession with an accuracy of about 10^{-5} and frame-dragging by the Earth spin with an accuracy of about 0.3% (ref. 47). According to the general theory of relativity, the direction marked by a gyro has a drift with respect to a direction defined by a ‘fixed star’, and so the drift of each gyroscope had to be measured with respect to a direction defined by a telescope pointing towards a distant guide star with known proper motion. The Gravity Probe B telescope pointed at the guide star

IM Pegasi (HR8703), a binary system, at a distance of about 300 light years from Earth; the proper motion of IM Pegasi with respect to distant quasars was then measured using VLBI with an accuracy of a fraction of a milliarcsecond per year. The Gravity Probe B telescope was designed to have a similar accuracy in pointing at the centre of the guide star (ref. 45 and see <http://einstein.stanford.edu/>).

If we spin an electrically charged sphere we produce a magnetic field. Any change of orientation of the sphere rotation axis would then generate a variation of magnetic field flux through a nearby circuit that would produce an induced electric current in the circuit. A simple example of these basic laws of electromagnetism is when we spin the magnet of the dynamo of our bicycle to power its light. The Gravity Probe B gyroscopes use magnetization induced by rotation of a superconductor^{48,49}. The magnetic moment of a superconductor rotating with respect to a local inertial frame is known as the London magnetic moment⁴⁹. For a spinning sphere the London moment is directed along the spin axis of the sphere. Each of the four Gravity Probe B gyroscopes consisted^{45,46} (see <http://einstein.stanford.edu/>) of a quartz sphere (rotor) of radius 1.9 cm, designed to be spherical and to have relative inhomogeneities of its density to a few parts in a million. Each rotor was covered with a very thin film of niobium, that is, a superconductor at the temperature of the experiment (about 2 K), and was spinning at approximately 4,000 r.p.m., so that the spinning superconductor layer generated a London magnetic moment aligned along the spin axis of the gyro. The rotors were encircled by a superconducting loop. The variations of the magnetic flux through the loop were measured by the changes of current in a superconducting quantum interference device (SQUID).

On 14 April 2007, after about 18 months of data analysis (see <http://einstein.stanford.edu/>), the first Gravity Probe B results were presented. The Gravity Probe B experiment measured the geodetic precession with an accuracy of the order of 1.5%. Indeed, the Gravity Probe B team discovered unexpected large drifts of the gyroscopes’ spin axes and estimated⁵⁰ the unmodelled systematic errors to be of the order of 100 milliarcseconds per year, corresponding to an uncertainty of the order of two and half times the frame-dragging by the Earth spin. However, by additional modelling of the systematic errors, the Gravity Probe B team aims to achieve (ref. 50 and see <http://einstein.stanford.edu/>) an uncertainty of about 5 milliarcseconds per year that would correspond to a measurement of the geodetic precession with about 0.1% accuracy and of frame-dragging by the Earth spin with about 13% accuracy.

The Gravity Probe B team has explained (ref. 51 and see <http://einstein.stanford.edu/>) the large drifts of the gyroscopes as being due to electrostatic patches on the surface of rotors and housings, producing unexpected classical torques on the gyroscopes and damping of their polhode motion, that is, the motion of the spin axis of a

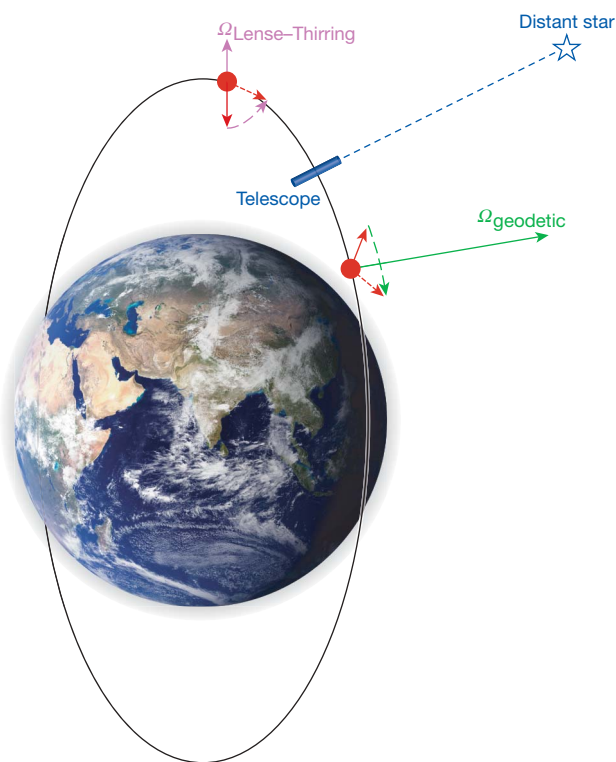


Figure 5 | Frame-dragging on the Gravity Probe B gyroscopes. The violet arrow displays frame-dragging of the Gravity Probe B gyroscopes (red arrows) by the Earth’s spin, Ω_{L-T} ; its average theoretical value is, under the general theory of relativity, about 39 milliarcseconds per year about an axis contained in the Gravity Probe B polar orbital plane. The green arrow represents the geodetic precession, Ω_{geodetic} ; its theoretical value is approximately 6,600 milliarcseconds per year about an axis orthogonal to the Gravity Probe B orbital plane (Box 1). (Earth’s image by NASA, <http://visibleearth.nasa.gov/>.)

gyroscope relative to its inertia ellipsoid. In addition to the well-known dynamics of a spinning toy top, we have a variety of astronomical and geophysical examples of the change of orientation of a body spin axis. A geophysical example of a torque-free spin axis motion is the Earth's polar motion or Chandler's wobble, that is, the wobbling of the Earth's rotation axis with a period of approximately 433 days. The precession of the equinoxes, that is, the precession of the Earth spin axis with a period of roughly 26,000 years, is an astronomical example of a torque-induced precession due to the effect of the gravitational fields of the Sun (and the Moon) on the non-spherical, flattened Earth. Thus, the external torques on the Gravity Probe B gyroscopes had to be reduced to a low-enough level to allow detection of the general relativistic shifts; however, this was not possible on the ground so the near-zero gravity conditions of an orbiting spacecraft were needed⁴⁶. Indeed, a drag-free (the Gravity Probe B non-gravitational accelerations were designed to be less than $10^{-10}g$), free-fall space experiment, with near-zero local gravity, should be free from most of the torques on the gyroscopes and should thus allow a measurement of the general relativistic drifts. Unfortunately, unexpected torques arose in space (see refs 50 and 51 and <http://einstein.stanford.edu/>; however, see also refs 52 and 53).

Lunar laser ranging and astrophysical observations

The general theory of relativity predicts that a gyroscope, at rest or moving with respect to a spinning mass, has a drag due to the spin of the mass, which is the Lense–Thirring effect on the LAGEOS satellites. If a gyroscope is at rest with respect to a non-rotating mass, it does not experience any drag. However, if the gyroscope starts to move with respect to the non-rotating mass it would acquire a precession that would again disappear when the gyroscope stops relative to the non-rotating mass. This precession of a gyroscope, dependent on its motion, is called geodetic precession or the de Sitter effect⁵⁴ (Box 1). The basic difference between frame-dragging by spin and geodetic precession is that in the case of the former (the Lense–Thirring effect) the rotation of a mass produces additional spacetime curvature, whereas in the case of the latter (the de Sitter effect) no spacetime curvature (see ref. 55, and section 6.11 of ref. 3) is produced by the motion of the test-gyroscope (for a discussion on frame-dragging and geodetic precession see refs 55–57, and section 6.11 of ref. 3). The geodetic precession has been accurately measured on the Moon's orbit^{58–60} and on Gravity Probe B and detected on binary pulsars^{61,62}.

Laser-ranging, to the Moon and to artificial satellites, is a basic tool for testing fundamental physics and general relativity^{15,23}. In 1969, a system of optical retro-reflectors was placed on the Moon's Mare Tranquillitatis (Sea of Tranquillity), by the Apollo 11 mission. Other retro-reflectors have been located on the Moon by the Apollo 14, Apollo 15, Luna 17 and Luna 21 missions. Using short laser pulses, the range from an emitting laser on Earth and a retro-reflector on the Moon is measured today with an accuracy of the order of a centimetre²³ that—for a mean Earth–Moon distance of approximately 385,000 km—corresponds to a fractional error in the distance of the order of a few parts in 10^{11} .

Lunar laser ranging had important applications for the study of the Moon's internal structure and of the dynamics of the Earth–Moon system. In fundamental physics, lunar laser ranging has provided very accurate tests of the equivalence principle, measurements of parameters testing the post-newtonian approximation of Einstein's gravitational theory, experimental limits on conceivable time variations of the gravitational constant and very accurate tests of the geodetic precession^{58–60}. The Moon's orbit around Earth is a kind of huge gyroscope, and so, because of the motion of the Earth–Moon system around the Sun, the node and perigee of the Moon have a geodetic precession with amplitude of about 19.2 milliarcseconds per year about an axis orthogonal to the ecliptic plane (Box 1); this shift is today measured⁵⁸ via lunar laser ranging with an uncertainty of approximately 0.6%. Similarly, by analysing the

frame-dragging effects due to the motion of the Earth–Moon system with respect to the Sun's mass, a test of frame-dependent gravitomagnetic effects on the Moon's orbit has been reported⁶³ with an accuracy of approximately 10^{-3} ; however, see different interpretations of this lunar laser ranging test in ref. 55, in section 6.11 of ref. 3, and in ref. 64. A detection by VLBI of the gravitomagnetic deflection of radio waves due to the orbital motion of Jupiter has also been reported⁶⁵, although different interpretations of this test are discussed in refs 66 and 9.

Outside the Solar System, frame-dragging effects have been observed in binary pulsars and on matter orbiting black holes and neutron stars, but the present detections of geodetic precession^{61,62} on binary pulsars and frame-dragging by spin on the accretion disks of black holes and neutron stars⁶⁷ have much lower accuracies than the corresponding measurements in the Solar System. Pulsars (rapidly rotating neutron stars) were discovered⁶⁸ in 1967, and are an outstanding tool for testing the general theory of relativity and fundamental physics⁶⁹. The received radio pulses correspond to the rotation of the neutron star, so pulsars provide stable astrophysical clocks. In particular, millisecond pulsars (with pulse periods of the order of milliseconds), discovered⁷⁰ in 1982, are extraordinarily accurate astrophysical clocks with long-term stabilities comparable to those of atomic clocks. In 1974, a binary pulsar, known as PSR B1913+16, was discovered⁷¹ (a binary pulsar consists of a pulsar in orbit with a companion star). Its pulse period is 59 milliseconds and its orbital period 7.75 hours. Since 1974, a number of binary pulsars have been discovered and they provide extraordinary astrophysical laboratories for testing the general theory of relativity via the measurement of their orbital parameters.

The geodetic precession of the spin axis of a binary pulsar, a spin-orbit frame-dragging effect, has been observed in the binary system PSR B1913+16 (ref. 61) and, in particular, in the system PSR B1534+12, where its measured value has been reported⁶² to be $0.44^{\circ} \left(\begin{smallmatrix} +0.48^{\circ} \\ -0.12^{\circ} \end{smallmatrix} \right)$ per year, in agreement with the general relativistic prediction of 0.52° per year. Using the double pulsar PSR J0737–3039A/B, a recently discovered unique binary system consisting of two radio pulsars, the geodetic precession could also be measured with improved accuracy⁷². Frame-dragging by spin may also be measured on binary pulsars. Astrophysical evidence of frame-dragging, due to the orbital mass currents by the two stars in the binary pulsar PSR B1913+16, is also provided by its periastron precession⁷³.

Despite all the observational tests passed by general relativity, we have recently discovered supernovae that appear to accelerate away from us^{74,75}. This can be explained by 'dark energy' that is accelerating the expansion of the Universe, and may imply a non-zero value to Einstein's 'cosmological constant', or an exotic new form of energy known as 'quintessence'^{76,77}, or perhaps a modification of the general theory of relativity. Together with dark matter, approximately 96% of the mass–energy of the Universe should be of a form unknown to us. A relativistic theory of gravity, combining the general theory of relativity at some level with quantum mechanics, may help to reveal what dark energy is and whether it might be related to dark matter.

New effects are being sought on very small length scales⁷⁸, on very large scales⁷⁹ and even in the Solar System⁸⁰ for what they might reveal about theories such as quantum gravity, string and brane-world models of the Universe. Furthermore, even though a breakdown of the general theory of relativity should occur at the quantum level, some viable modifications of Einstein's theory already give different predictions at the classical level⁸. I thus believe that every aspect of the general theory of relativity should be directly tested and the accuracy of the present measurements of Einstein's theory and of the foundations of gravitational theories should be further improved. The planned space observatories will open new windows on the Universe, and I expect surprises that may change our way of thinking about the Universe.

1. Misner, C. W., Thorne, K. S. & Wheeler, J. A. *Gravitation* (Freeman, San Francisco, 1973).
2. Weinberg, S. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972).
3. Ciufolini, I. & Wheeler, J. A. *Gravitation and Inertia* (Princeton Univ. Press, Princeton, New Jersey, 1995).
4. Einstein, A. Letter to Ernst Mach. Zurich, 25 June 1913. In *Gravitation* (Misner, C. W., Thorne, K. S. & Wheeler, J. A.) 544 (Freeman, San Francisco, 1973).
5. Mach, E. *Die Mechanik in Ihrer Entwicklung Historisch-Kritisch-Dargestellt* (Brockhaus, Leipzig, 1912); transl. *The Science of Mechanics* (Open Court, La Salle, Illinois, 1960).
6. Barbour, J. & Pfister, H. (eds) *Mach's Principle. From Newton's Bucket to Quantum Gravity* (Birkhauser, Boston, 1995).
7. Will, C. M. *Was Einstein Right? Putting General Relativity to the Test* 2nd edn (Basic Books, New York, 1993).
8. Will, C. M. *Theory and Experiment in Gravitational Physics* 2nd edn (Cambridge Univ. Press, Cambridge, UK, 1993).
9. Will, C. M. The confrontation between general relativity and experiment. *Living Rev. Rel.* **9**, 3 (2006); (<http://www.livingreviews.org/lrr-2006-3>).
10. Landau, L. D. & Lifshitz, E. M. *The Classical Theory of Fields* 3rd rev., English edn. (Pergamon, London, 1971).
11. Zeldovich, YaB & Novikov, I. D. *Relativistic Astrophysics Vol. I Stars and Relativity* (Univ. Chicago Press, Chicago, 1971).
12. Ciufolini, I. & Ricci, F. Time delay due to spin inside a rotating shell. *Class. Quantum Grav.* **19**, 3875–3881 (2002).
13. Thorne, K. S., Price, R. H. & Macdonald, D. A. *The Membrane Paradigm* (Yale Univ. Press, New Haven, 1986).
14. Schäfer, G. Gravitomagnetic effects. *J. Gen. Rel. Grav.* **36**, 2223–2235 (2004).
15. Ciufolini, I. & Pavlis, E. C. A confirmation of the general relativistic prediction of the Lense–Thirring effect. *Nature* **431**, 958–960 (2004).
16. Ciufolini, I., Pavlis, E. C. & Peron, R. Determination of frame-dragging using Earth gravity models from CHAMP and GRACE. *New Astron.* **11**, 527–550 (2006).
17. Ashtekar, A. Gravity, geometry and the quantum. In *Proc. Albert Einstein Century Int. Conf. (Paris, 18–22 July 2005)* (eds Alimi, J.-M. & Füzfa, A.) (American Institute of Physics, New York, 2006).
18. Penrose, R. Gravitational collapse and space–time singularities. *Phys. Rev. Lett.* **14**, 57–59 (1965).
19. Hawking, S. W. & Ellis, G. F. R. *The Large Structure of Space-time* (Cambridge Univ. Press, Cambridge, 1973).
20. de Sitter, W. On Einstein's Theory of Gravitation and its astronomical consequences. *Mon. Not. R. Astron. Soc.* **77**, 155–184 (1916).
21. Lense, J. & Thirring, H. Über den Einfluss der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie. *Phys. Z.* **19**, 156–163 (1918).
22. Bender, P. L. et al. The lunar laser ranging experiment. *Science* **182**, 229–238 (1973).
23. Williams, J. G., Turyshev, S. G. & Murphy, T. W. Jr. Improving LLR tests of gravitational theory. *Int. J. Mod. Phys. D* **13**, 567–582 (2004).
24. Cohen, S. C. & Dunn, P. J. LAGEOS scientific results. *J. Geophys. Res.* **B 90**, 9215–9438 (1985).
25. Noomen, R., Klosko, S., Noll, C. & Pearlman, M. (eds) *Toward Millimeter Accuracy. NASA CP 2003–212248 Proc. 13th Int. Laser Ranging Workshop* (NASA Goddard, Greenbelt, Maryland, 2003).
26. Ciufolini, I. Measurement of the Lense–Thirring drag on high-altitude laser-ranged artificial satellites. *Phys. Rev. Lett.* **56**, 278–281 (1986).
27. Ciufolini, I. On a new method to measure the gravitomagnetic field using two orbiting satellites. *Nuovo Cimento. A* **109**, 1709–1720 (1996).
28. Ciufolini, I. A comprehensive introduction to the LAGEOS gravitomagnetic experiment. *Int. J. Mod. Phys. A* **4**, 3083–3145 (1989).
29. Ries, J. C., Eanes, R. J., Watkins, M. M. & Tapley, B. *NASA-ASI Study, Report Part A* (CSR, Austin, 1989).
30. Ciufolini, I. et al. *Italian Space Agency Phase A Report on LARES*. (Italian Space Agency, ASI, Rome, 1998).
31. Rubincam, D. P. Drag on the LAGEOS satellite. *J. Geophys. Res.* **B 95**, 4881–4886 (1990).
32. Lucchesi, D. M. Reassessment of the error modelling of non-gravitational perturbations on LAGEOS II and their impact in the Lense–Thirring derivation. Part II. *Planet. Space Sci.* **50**, 1067–1100 (2002).
33. Kaula, W. M. *Theory of Satellite Geodesy* (Blaisdell, Waltham, 1966).
34. Reigber, C. et al. An Earth gravity field model complete to degree and order 150 from GRACE: EIGENGRACE02S. *J. Geodyn.* **39**, 1–10 (2005).
35. Ciufolini, I., Pavlis, E. C., Chieppa, F., Fernandes-Vieira, E. & Perez-Mercader, J. Test of general relativity and measurement of the Lense–Thirring effect with two Earth satellites. *Science* **279**, 2100–2103 (1998).
36. Ciufolini, I., Chieppa, F., Lucchesi, D. & Vespe, F. Test of Lense–Thirring orbital shift due to spin. *Class. Quantum Grav.* **14**, 2701–2726 (1997).
37. Ries, J. C., Eanes, R. J., Tapley, B. D. & Peterson, G. E. Prospects for an improved Lense–Thirring test with SLR and the GRACE gravity mission. In *Toward Millimeter Accuracy. Proc. 13th Int. Laser Ranging Workshop* (eds Noomen, R., Klosko, S., Noll, C. & Pearlman, M.) Report NASA CP 2003–212248 (NASA Goddard, Greenbelt, Maryland, 2003).
38. Pavlis, E. C. in *Recent Developments in General Relativity (Genoa, 2000)* (eds Cianci, R. et al.) 217–233 (Springer Italia, Milan, 2002).
39. Reigber, C. et al. GRACE orbit and gravity field recovery at GFZ Potsdam—first experiences and perspectives. *Eos (Fall Meet. Suppl.)* **83** (47), abstr. G12B–03 (2002).
40. Tapley, B. D. The GRACE mission: status and performance assessment. *Eos (Fall Meet. Suppl.)* **83** (47), abstr. G12B–01 (2002).
41. Ciufolini, I. et al. in *Proc. First Intl School of Astrophysical Relativity “John Archibald Wheeler”: Frame-Dragging, Gravitational-Waves and Gravitational Tests* (eds Ciufolini, I. & Matzner, R. A.) (Springer, in the press).
42. Pugh, G. E. *Proposal for a Satellite Test of the Coriolis Prediction of General Relativity*. Weapons Systems Evaluation Group Research Memorandum N. 11 (The Pentagon, Washington, 1959).
43. Schiff, L. I. Motion of a gyroscope according to Einstein's Theory of Gravitation. *Proc. Natl. Acad. Sci.* **46**, 871–882 (1960).
44. Schiff, L. I. Possible new test of General Relativity Theory. *Phys. Rev. Lett.* **4**, 215–217 (1960).
45. Everitt, C. W. F. et al. Papers on the Stanford Relativity gyroscope experiment (NASA Gravity Probe B). *Proc. SPIE*, **619**, 29–165 (Society for Photo-Optical Instrumentation Engineers, Washington, 1986).
46. Fitch, V. L. et al. *Review of Gravity Probe B* (National Academic Press, Washington DC, 1995).
47. Buchman, S., Everitt, C. W. F., Parkinson, B., Turneaure, J. P. & Keiser, G. M. Cryogenic gyroscopes for the relativity mission. *Physica B* **280**, 497–498 (2000).
48. Becker, R., Heller, G. & Sauter, F. Über die Stromverteilung in Einer Supraleitenden Kugel. *Z. Phys.* **85**, 772–787 (1933).
49. London, F. *Superfluids Vol. 1* (Wiley, New York, 1950).
50. Muhlfelder, B., Keiser, G. M. & Turneaure, J. *GP-B Experiment Error: A Work in Progress*. (Stanford Univ., Stanford, April 2007); poster at (http://einstein.stanford.edu/content/aps_posters/ExperimentError.pdf).
51. Gill, D. K. & Buchman, S. *Evidence for Patch Effect Forces On the Gravity Probe B Gyroscopes*. (Stanford Univ., Stanford, April 2007); poster at (http://einstein.stanford.edu/content/aps_posters/EvidenceForPatchEffectForces.pdf).
52. Barker, B. M. & O'Connell, R. F. The gyroscope test of General Relativity. *Nature* **312**, 314 (1984).
53. Barker, B. M. & O'Connell, R. F. The gravitational interaction: spin, rotation, and quantum effects—a review. *Gen. Rel. Grav.* **11**, 149–175 (1979).
54. de Sitter, W. On Einstein's Theory of Gravitation and its Astronomical Consequences. *Mon. Not. R. Astron. Soc.* **76**, 699–728 (1916).
55. Ciufolini, I. Gravitomagnetism, frame-dragging and lunar laser ranging. Preprint at (<http://xxx.lanl.gov/abs/0704.3338>) (10 May, 2007).
56. Ashby, N. & Shahid-Saless, B. Geodetic precession or dragging of inertial frames? *Phys. Rev. D* **42**, 1118–1122 (1990).
57. O'Connell, R. F. A note on frame dragging. *Class. Quant. Grav.* **22**, 3815–3816 (2005).
58. Williams, J. G., Turyshev, S. G. & Boggs, D. H. Progress in lunar laser ranging tests of relativistic gravity. *Phys. Rev. Lett.* **93**, 261101 (2004).
59. Williams, J. G., Newhall, X. X. & Dickey, J. O. Relativity parameters determined from lunar laser ranging. *Phys. Rev. D* **53**, 6730–6739 (1996).
60. Bertotti, B., Ciufolini, I. & Bender, P. L. New test of general relativity: measurement of de Sitter geodetic precession rate for lunar perigee. *Phys. Rev. Lett.* **58**, 1062–1065 (1987).
61. Weisberg, J. M. & Taylor, J. H. General relativistic geodetic spin precession in binary pulsar B1913+16: mapping the emission beam in two dimensions. *Astrophys. J.* **576**, 942–949 (2002).
62. Stairs, I. H., Thorsett, S. E. & Arzoumanian, Z. Measurement of gravitational spin-orbit coupling in a binary-pulsar system. *Phys. Rev. Lett.* **93**, 141101 (2004).
63. Murphy, T. W. Jr, Nordtvedt, K. & Turyshev, S. G. Gravitomagnetic influence on gyroscopes and on the lunar orbit. *Phys. Rev. Lett.* **98**, 071102 (2007).
64. Kopeikin, S. M. Comment on “The gravitomagnetic influence on gyroscopes and on the lunar orbit”. *Phys. Rev. Lett.* **98**, 229001 (2007).
65. Kopeikin, S. M. & Fomalont, E. B. Gravitomagnetism, causality, and aberration of gravity in the gravitational light-ray deflection experiments. *Gen. Rel. Grav.* (in the press); preprint at (<http://xxx.lanl.gov/abs/gr-qc/0510077>) (2005).
66. Will, C. M. Propagation speed of gravity and the relativistic time delay. *Astrophys. J.* **590**, 683–690 (2003).
67. Cui, W. et al. Evidence for frame-dragging around spinning black holes in X-ray binaries. *Astrophys. J.* **492**, L53–L58 (1998).
68. Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F. & Collins, R. A. Observation of a rapidly pulsating radio source. *Nature* **217**, 709–713 (1968).
69. Stairs, I. H. Testing General Relativity with pulsar timing. *Living Rev. Rel.* **6**, 5 (2003); (<http://relativity.livingreviews.org/articles/lrr-2003-5/index.html>).
70. Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M. & Goss, W. M. A millisecond pulsar. *Nature* **300**, 615–618 (1982).
71. Hulse, R. A. & Taylor, J. H. Discovery of a pulsar in a binary system. *Astrophys. J.* **195**, L51–L53 (1975).
72. Kramer, M. et al. Tests of general relativity from timing the double pulsar. *Science* **314**, 97–102 (2006).

73. Nordtvedt, K. Existence of the gravitomagnetic interaction. *Int. J. Theor. Phys.* **27**, 1395–1404 (1988).
74. Riess, A. *et al.* Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.* **116**, 1009–1038 (1998).
75. Perlmutter, S. *et al.* Measurements of Ω and Λ from 42 high-redshift supernovae. *Astrophys. J.* **517**, 565–586 (1999).
76. Perlmutter, S. Supernovae, dark energy, and the accelerating universe. *Phys. Today* **56**, 53–59 (2003).
77. Caldwell, R. R. Dark energy. *Phys. World* **17**, 37–42 (2004).
78. Adelberger, E., Heckel, B. & Hoyle, C. D. Testing the gravitational inverse-square law. *Phys. World* **18**, 41–45 (2005).
79. Amelino-Camelia, G., Ellis, J., Mavromatos, N. E., Nanopoulos, D. V. & Sarkar, S. Potential sensitivity of gamma-ray burster observations to wave dispersion *in vacuo*. *Nature* **393**, 763–765 (1998).
80. Dvali, G. Filtering gravity: modification at large distances? In *Proc. Nobel Symp. 127 on String Theory and Cosmology (Sigtuna, Sweden, 2003)* (eds Danielsson, U., Goobar, A. & Nilsson, B.) (World Scientific, Singapore, 2005); also preprint at (<http://arXiv.org/hep-th/0402130>) (2004).
81. Kerr, R. P. Gravitational field of a spinning mass as an example of algebraically special metrics. *Phys. Rev. Lett.* **11**, 237–238 (1963).
82. Ciufolini, I. *et al.* *NASA-ASI Study, Report Part B* (ASI, Rome, 1989).

Acknowledgements I gratefully acknowledge the support of the Italian Space Agency and the comments of R. Matzner of the Centre for Relativity, University of Texas at Austin, I. Novikov of the Astro Space Centre, Lebedev Physical Institute Moscow and D. P. Rubincam of NASA-Goddard.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to I.C. (ignazio.ciufolini@unile.it).

ARTICLES

An asteroid breakup 160 Myr ago as the probable source of the K/T impactor

William F. Bottke¹, David Vokrouhlický^{1,2} & David Nesvorný¹

The terrestrial and lunar cratering rate is often assumed to have been nearly constant over the past 3 Gyr. Different lines of evidence, however, suggest that the impact flux from kilometre-sized bodies increased by at least a factor of two over the long-term average during the past ~100 Myr. Here we argue that this apparent surge was triggered by the catastrophic disruption of the parent body of the asteroid Baptistina, which we infer was a ~170-km-diameter body (carbonaceous-chondrite-like) that broke up 160^{+30}_{-20} Myr ago in the inner main asteroid belt. Fragments produced by the collision were slowly delivered by dynamical processes to orbits where they could strike the terrestrial planets. We find that this asteroid shower is the most likely source (>90 per cent probability) of the Chicxulub impactor that produced the Cretaceous/Tertiary (K/T) mass extinction event 65 Myr ago.

The nature of the terrestrial and lunar impact flux over the past several Gyr has been subject to considerable debate, with key arguments focusing on the dominance of asteroidal versus cometary impactors and whether the flux is constant, cyclic, or is punctuated in some manner with random ‘showers’. Although all attempts to identify a cycle period in the impact crater data have largely failed, the issue of whether the impact flux has varied, and if so by how much, over what timescale, and by what mechanism, is nevertheless an important one for understanding the geological and biological history of our planet. The fact that the current population of near-Earth objects (NEOs) looks reasonably close to what would be expected from a constant background flux from the asteroid belt¹, and that the present-day impact flux from those bodies seems to match up, more or less, with the long-term average flux over billions of years^{2–4}, suggests that any putative showers must be either of very short duration or of modest increase over the background for a more prolonged period.

In this light, it is useful to examine the recent impact history of the Earth and Moon to see whether insights into these issues can be gleaned from the best preserved and most accessible craters. An analysis of terrestrial craters found on stable cratons in North America, Europe and Australia suggests the collision rate of diameter $D \geq 1$ km projectiles may have increased by a factor of ~2 or so over the past ~100 Myr (refs 3–6). Some even argue that the terrestrial crater record supports a fourfold increase over this time, though this has yet to be verified⁷. Studies of lunar craters also support a factor of 2 increase in the recent impact flux (for example, counts of craters found within $D > 8$ km craters on the lunar nearside⁸, counts of craters on/near the 109-Myr-old Tycho crater⁴, and estimates of the impact flux from lunar rayed craters⁵, although also see ref. 6). These data sets should be interpreted with some caution; we note that few craters on the Moon have well-determined ages. Still, a recent increase in the multi-kilometre impactor flux would provide the simplest way to explain the abundance of impact-derived glass spherules with relatively young ages found in lunar soils (see, for example, ref. 9).

In the past, it was assumed that sudden changes in the terrestrial planet impactor flux must come from comet showers, presumably produced by the passage of a star through the Oort cloud (see, for example, ref. 3). The expected duration of these showers, however, is only a few Myr (ref. 10), too short to explain the crater and impact-spherule age distributions described above. Moreover, our

current understanding of the active and dormant populations of Jupiter-family comets^{2,11} and nearly isotropic comets^{11,12} (see also Supplementary Discussion) suggests that these objects strike the Earth and Moon too infrequently to be considered a plausible source for any prolonged increase in the impact flux. The comet shower model has been further weakened by the fact that several large terrestrial craters once assumed to have been produced by comets³ have now been linked to projectiles with asteroid-like compositions^{13,14}.

We instead argue here that the increased impact flux described above was set off by the catastrophic disruption of a ~170-km body in the innermost region of the main asteroid belt. This breakup event produced what is now known as the Baptistina asteroid family (BAF), a cluster of fragments with similar proper semimajor axes a , eccentricities e and inclinations i . According to our numerical simulations, the BAF’s location, age and fragment size distribution are remarkably well suited to generate a ~100-Myr-long surge in the multi-kilometre NEO population and explain the above observations. Moreover, the bombardment created by the BAF provides the most probable source for the projectiles that produced the K/T impact on Earth and the conspicuous Tycho crater on the Moon. Interpreted more broadly, this event can be considered powerful evidence for a recurrent process in which the formation of inner main belt families like Flora or Vesta leads to showers of debris delivered to Earth^{15–17}.

Identifying the Baptistina family

The BAF is centred on (298) Baptistina, a $D \approx 40$ km asteroid with proper semimajor axis $a_{\text{bap}} = 2.26$ AU (ref. 18) and spectroscopic characteristics similar to carbonaceous chondrites (that is, those resembling C- or X-type asteroids)¹⁹ (see Supplementary Discussion). Figure 1 shows our best estimate of the observed BAF and also describes how it was identified. The family resides in a complicated region of the main belt; it is partially overlapped by the prominent Flora and Vesta families, and is nearly bisected by the 7:2 and 5:9 mean motion resonances with Jupiter and Mars, respectively²⁰ (we refer to these mean motion resonances together as the J7:2/M5:9).

To pick out BAF members from the background, we take advantage of recent advances in our knowledge of the dynamical evolution and spectroscopic properties of asteroid families. For the former, it is now recognized that small asteroids are affected by the Yarkovsky and YORP (Yarkovsky–O’Keefe–Radzievskii–Paddack) effects, thermal

¹Southwest Research Institute, 1050 Walnut St, Suite 300, Boulder, Colorado 80302, USA. ²Institute of Astronomy, Charles University, V Holešovičkách 2, 18000 Prague 8, Czech Republic.

radiation forces and torques that cause $D < 40$ km objects to undergo semimajor axis drift and spin vector modifications, respectively, as a function of their size, spin vector, orbit and physical/material properties²¹. These non-gravitational perturbations leave their mark on families, causing them to evolve in ways that produce characteristic fingerprints in orbital element space²² (see Supplementary Discussion).

Immediately after an asteroid breakup event, individual fragments begin to spread in semimajor axis a by the Yarkovsky effect, with smaller bodies drifting inward towards or outward away from the Sun faster than larger bodies²¹. At the same time, the YORP effect works to preferentially tilt the obliquities of the fragments towards 0° or 180° , values that optimize their da/dt drift rates. Numerical simulations show that over time, the coupled Yarkovsky/YORP effects cause a family to evolve into a two-lobed structure in a and absolute magnitude H space (a, H), with each lobe filled with fast-moving asteroids headed for the family's periphery²² (see Supplementary Discussion). This dynamical signature, which has already been identified in several young families²², is seen in Fig. 1. The dark grey lines show our estimate of the borders of the BAF. According to Yarkovsky/YORP evolution models, the BAF boundaries are defined as the outermost region of each lobe in (a, H) space where the number density of asteroids is significantly higher than in the background. Objects outside the lines should mostly be background objects (see Supplementary Discussion).

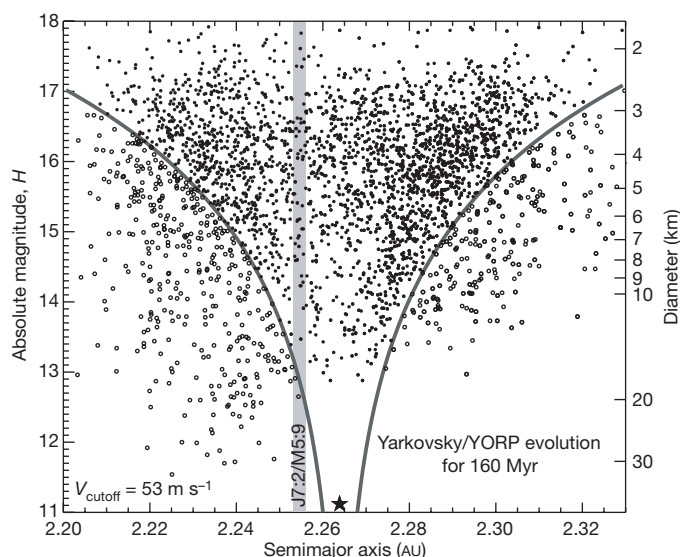


Figure 1 | The orbital and size distribution of the observed Baptistina asteroid family. The family has been projected onto a plane of proper semimajor axis a versus absolute magnitude H . On the right ordinate, we show asteroid diameters for a typical C-type asteroid albedo of 0.04. The central and largest body of the family, C-type asteroid (298) Baptistina¹⁹, has proper semimajor axis $a = 2.264$ AU, eccentricity $e = 0.15$, and sine of inclination $\sin i = 0.10$ (ref. 18). The BAF was identified using the hierarchical clustering method (HCM) applied to the proper orbital element database found in the AstDyS database¹⁸ (see Supplementary Discussion). The HCM locates bodies in the neighbourhood of (298) Baptistina with mutual velocities less than a threshold limit V_{cutoff} . The filled and open circles show 3,042 linked objects with $V_{\text{cutoff}} = 53 \text{ m s}^{-1}$. The family is also noticeably depleted near the adjacent J7:2 and M5:9 resonances²⁰ (grey bar at $a \approx 2.2545$ AU). The two-lobed structure with an evacuated centre is diagnostic of families that have spread in a for an extended time under the influence of Yarkovsky/YORP thermal forces^{21,22}. The dark grey lines that bracket the outside of each lobe represent our best estimate of how far the majority of family members could have spread in ~ 160 Myr. Objects outside these curves, shown as open circles, are assumed to be predominantly interlopers. Most come from the nearby Flora or Vesta families, whose spectroscopic signatures are similar to space-weathered ordinary chondrites (that is, S-type asteroids; Flora family) or basaltic achondrites (HED) meteorites (that is, V-type asteroids; Vesta family)^{19,43,45}. Their number density near a H of ~ 16 , at which the SDSS is sensitive to asteroids, indicates that interlopers between the grey curves only contribute 10–20% to the overall BAF.

To test our BAF identification scheme, we examined the colour properties of BAF members and background asteroids in five photometric bands using data from the Sloan Digital Sky Survey (SDSS)²³. For values near a H of ~ 16 , the number of asteroids with C- or X-type colours between the grey lines in Fig. 1 were found to dominate those with S- or V-type-like colours by at least a factor of 4, values consistent with our estimates (see Supplementary Discussion). Our analysis also showed that all of the asteroids with C- or X-type colours in the innermost region of the main belt near the BAF reside between the grey lines. This provides support for the idea that (298) Baptistina and our predicted BAF members came from the same disruption event.

Size distribution and parent body size

An intriguing aspect of Fig. 1 is the apparent deficiency of BAF members found close to the J7:2/M5:9 relative to the population found near its 'mirror image' location at ~ 2.27 AU, with the centre of symmetry defined as the semimajor axis of (298) Baptistina. Numerical simulations indicate that bodies in the J7:2/M5:9 can have their eccentricities pumped up to Mars-crossing orbits over timescales as short as a few Myr (ref. 20). We infer that an unknown fraction of the BAF has escaped the main belt and entered the terrestrial planet region since the family-forming event took place. To quantify how many kilometre-sized and larger fragments have been lost over time, we must estimate the size–frequency distribution (SFD) created by the BAF's formation and how it has changed over the family's lifetime.

Yarkovsky/YORP modelling work indicates that families with two-lobed structures like those in Fig. 1 are essentially symmetric in (a, H) space²². This means that if the J7:2/M5:9 did not exist, the population on the left side of the BAF should reflect that of the right side. Hence, by doubling the population of the right side of the family (Fig. 1), we obtain an estimate of the BAF's original SFD that accounts for the missing objects on the left side, provided that collisional evolution is unimportant. This 'derived SFD' is shown in Fig. 2.

The cumulative power-law index q of our derived SFD between $10 \text{ km} < D < 20 \text{ km}$ (cumulative number $N(>D) \propto D^q$) is -6.5 to -8 , steep enough to produce ~ 300 $D > 10$ km objects. It then bends to $q \approx -2.7$ for $6 \text{ km} < D < 10 \text{ km}$. Using a Monte Carlo code to account for observational and statistical errors in the H values of known BAF members (see Supplementary Discussion), we fit a power law to the distribution of $6 \text{ km} < D < 10 \text{ km}$ objects, a size range where we are confident the population is both observationally complete and has been relatively unaffected by collisional evolution. This power law was then extrapolated down to $D = 1 \text{ km}$. According to our best-fit results, the initial BAF once contained $(1.36 \pm 0.3) \times 10^5$ objects with $D > 1 \text{ km}$.

The size of the Baptistina parent body cannot be directly computed from observations, partly because $D \lesssim 3 \text{ km}$ BAF members are too small to be detected by existing surveys but also because many have been removed by collisional and dynamical processes. To circumvent this problem, we used numerical impact experiments to gain insights into the circumstances that formed the BAF²⁴. Our simulations, created using a smooth particle hydrodynamic (SPH) code coupled to an N -body code, followed projectiles shot into coherent target asteroids over a wide range of projectile/target mass ratios, collision velocities, and impact angles²⁵. The resultant model SFDs were then compared to the largest bodies of the derived SFD described above; these objects have yet to experience significant collisional evolution²⁴. Because the model SFD must explicitly conserve mass down to the code's resolution limit, a reasonable match at the large diameter end of the SFD allows us to estimate how much of the BAF's mass was initially in the form of smaller objects²⁴. Figure 2 shows our most successful run to date. It indicates that the Baptistina parent body had $D \approx 170 \text{ km}$, and that the BAF's initial SFD had 88% of its mass in the form of $D < 10 \text{ km}$ fragments.

Collisional evolution

Using the derived SFD together with the estimated size of the Baptistina parent body, we computed how collisional evolution has

affected the BAF's SFD over time. This was done using CoDDEM, a one-dimensional self-consistent code capable of tracking the collisional evolution and dynamical depletion of both the BAF and main belt SFDs simultaneously^{1,26}. The starting SFD for the BAF was chosen to match the derived SFD in Fig. 2 as well as constraints provided by the steep fresh crater population on nearby asteroid (951) Gaspra²⁷. The total mass represented by the SFD was required to match the estimated mass of the Baptistina parent body. The main belt SFD was assumed to be in a quasi-steady state, where small objects removed by collisions or the Yarkovsky effect are continually replenished by fresh fragments produced by larger breakup events¹.

Our results show that the BAF's SFD quickly grinds itself down to take on the same shape as the background main belt SFD for $D < 5$ km objects (Fig. 3). This explains why small Baptistina fragments do not dominate the inner main belt population today. Among larger objects, we found that 200 Myr after the BAF formation event took place, the number of $D > 1$ km family members had decreased by $\sim 40\%$. Accordingly, unless the BAF is extremely young, the number of km-sized fragments escaping the main belt today via the J7:2/M5:9 must be a factor of ~ 2 lower than the number that escaped immediately after the BAF formation event.

Dynamical evolution

Additional insights into the BAF's age and dynamical history can be gleaned by modelling how the observed family members have evolved from their initial orbits to those in Fig. 1. This was done by numerically

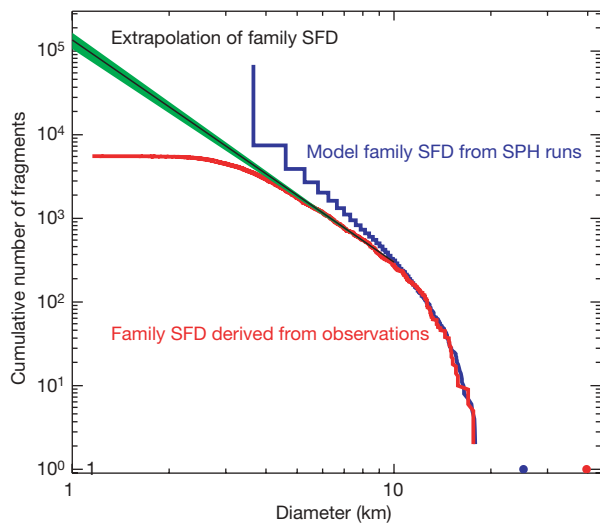


Figure 2 | Estimates of the initial size-frequency distribution of the Baptistina family. The red 'derived SFD' comprises (298) Baptistina and twice the family SFD with $a > a_{\text{hap}}$ (4,839 objects in all). The rollover on the left is caused by observational incompleteness. The black curve shows a best fit power law extrapolation from the distribution of $6 \text{ km} < D < 10 \text{ km}$ objects to $D = 1 \text{ km}$, whereas the green shading shows the standard deviation in the fit (see Supplementary Discussion). The blue SFD shows the best fit between the red SFD and model SFDs taken from a suite of 161 SPH/N-body impact simulations²⁵. The target asteroids were $D = 100 \text{ km}$ solid basalt spheres made up of 100,000 individual SPH components. The projectiles were given $D = 10\text{--}46 \text{ km}$, impact speeds of $2.5\text{--}7 \text{ km s}^{-1}$ and impact angles of $15^\circ\text{--}75^\circ$ (that is, nearly head-on to very oblique). The results of each simulation were scaled and compared to the red SFD to estimate the true size of the parent body²⁴. The match between the red and blue curves is very good for $10 \text{ km} < D < 20 \text{ km}$ objects. Although there is a mismatch between the largest remnants of each SFD, which differ in diameter by a factor of 1.6, these objects comprise less than 1% of the parent body's total mass. At $D < 10 \text{ km}$, the SFDs diverge as the model SFD approaches its resolution limit (that is, all model fragments must be in the form of $D > 3.5 \text{ km}$ bodies; this effectively creates an overabundance of $3.5 \text{ km} < D < 10 \text{ km}$ bodies). With considerable parameter space left to explore, our best fit run indicates that the BAF was created by the impact of a $\sim 60\text{-km}$ projectile striking nearly vertical to the surface of a $D \approx 170 \text{ km}$ target body at 3 km s^{-1} .

tracking the simultaneous dynamical and spin vector evolution of test asteroids affected by both the Yarkovsky and YORP effects²². Details of our code and model parameters can be found in the Supplementary Discussion. Our test asteroids, which were given the same sizes as observed BAF members (Fig. 1), were assumed to have been ejected from the current orbit of (298) Baptistina with random trajectories and velocities $V = V_0(D_0/D)$, with $D_0 = 5 \text{ km}$, D the diameter of our test asteroids in km, and V_0 a solved-for velocity parameter in our code. From this initial orbital distribution, the test asteroids were allowed to drift in a according to an established formulation of Yarkovsky/YORP evolution. The bulk density of our test asteroids was set to 1.3 g cm^{-3} , the average bulk density for C-type asteroids²⁸. The goal of our runs was to reproduce the characteristic signature of the BAF in (a, H) , with variables being time since the family-forming event, V_0 , and a parameter C that controls the strength of the YORP effect.

The best fit between our model population and the observed BAF yields a formation age of 160^{+30}_{-20} Myr and characteristic ejecta velocity dispersion $V_0 = 40 \pm 10 \text{ m s}^{-1}$. This makes the BAF the youngest family of its size in the innermost region of the main belt. The timing of the BAF breakup is also interesting, in that it raises the possibility that the Baptistina family-forming event could have influenced the terrestrial planet impactor flux over the Cretaceous period (65–145 Myr ago) and beyond. This would have been accomplished by large BAF members being directly injected and/or driven by the

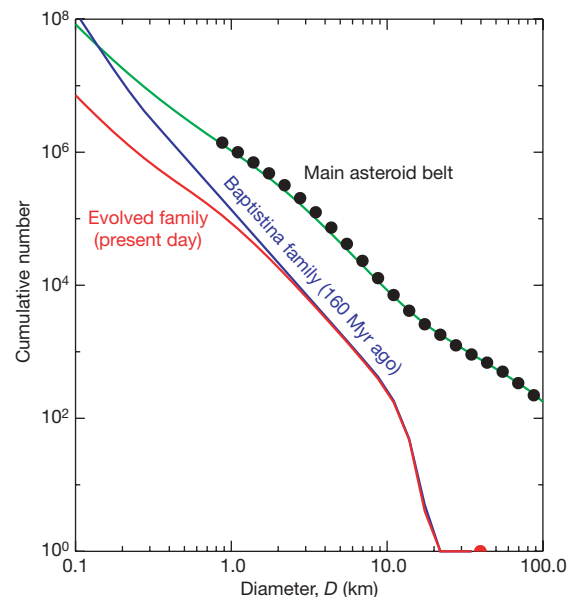


Figure 3 | Collisional evolution of the Baptistina family's size distribution. Using the code CoDDEM^{1,26}, we tracked the evolution of the BAF and main belt SFDs, assuming the objects in each population had bulk densities of 1.3 and 2.7 g cm^{-3} , respectively²⁸. The main belt SFD, asteroid disruption law, and fragment SFDs produced by each individual breakup event are described in ref. 1. The filled circles represent the main belt SFD derived from observations, whereas the green curve is our model main belt SFD. We determined that the intrinsic collision probability (P_i) of Baptistina fragments with one another and with main belt asteroids was 18×10^{-18} and $2.8 \times 10^{-18} \text{ km}^{-2} \text{ yr}^{-1}$, respectively, whereas the impact velocities V_{imp} were 3.9 and 5 km s^{-1} , respectively⁵⁰. The SFDs were tracked for 200 Myr. The figure shows the evolution of the BAF's SFD, assuming that the family forming event took place 160 Myr ago (blue curve). We see that the shape of the BAF's evolved SFD mimics that of the background main belt population for $D < 5 \text{ km}$ objects (red curve). This is because smaller objects are eliminated faster than they can be replenished by disruption events among bigger objects. Conversely, $D > 5 \text{ km}$ objects are harder to disrupt, making them relatively immune to collisional evolution over the timescale of our simulation. Overall, we find the population of $D > 1 \text{ km}$ fragments roughly decreases by 10% over 10 Myr, 20% over 40 Myr, 30% over 100 Myr and 40% over 200 Myr (see Supplementary Discussion).

Yarkovsky effect into the nearby J7:2/M5:9, which is capable of pumping up their eccentricities to planet-crossing values²⁰. Although most BAF escapees would ultimately be eliminated by hitting the Sun or being ejected out of the inner Solar System by an encounter with Jupiter, a small fraction would have gone on to strike the Earth or other planets^{2,15}.

The length and severity of an asteroid shower from the BAF is a function of the BAF's SFD, the likelihood of BAF members reaching the J7:2/M5:9, and the efficiency of the J7:2/M5:9 at producing planetary impactors. To determine the fraction of BAF members pushed out of the main belt by the J7:2/M5:9, we created a model of the initial orbital distribution of the family immediately after the breakup event. Here we took the largest 800 BAF objects with $a > a_{\text{bap}}$ found in Fig. 1 and gave them starting values (a_{start}) consistent with the BAF's initial ejection velocity distribution; $a_{\text{bap}} + (a - a_{\text{bap}})/2$ and $a_{\text{bap}} - (a - a_{\text{bap}})/2$. Next, we tracked the dynamical evolution of test family members using the symplectic integration code SWIFT-RMVS3 (ref. 29) modified to accommodate Yarkovsky thermal forces³⁰. The bodies were assigned sizes of $D = 1, 5$ and 10 km, bulk densities of 1.3 g cm^{-3} (ref. 28), spin periods of 6 ± 2 h (ref. 31), random spin axis orientations, and a thermal inertia of $100 \text{ J m}^{-2} \text{ s}^{-0.5} \text{ K}^{-1}$ (ref. 32). We tracked 1,600 bodies with $D = 1$ km and 900 bodies with $D = 5$ and 10 km. The planets Venus through to Neptune were included in our integrations. See Supplementary Discussion for example runs.

Overall, about 20% of our $D = 5$ and 10 km model asteroids reached the J7:2/M5:9 and escaped the main belt over 160 Myr (see Fig. 4 legend). For our $D = 1$ km bodies, we scaled our raw integration results to account for the effects of YORP; over 10–20 Myr timescales, YORP can spin up a km-sized body to the point that it sheds mass, or spin it down to the point that it readily undergoes a spin axis reorientation event³³ (see also ref. 21). We estimate that these so-called YORP cycles slow the nominal Yarkovsky drift rates of $D = 1$ km bodies by a factor of ~ 3 or so, such that only 13% escaped after 160 Myr.

These runs were also used to determine how many BAF members were in the NEO population today. By tracking the number of model asteroids in the NEO population at 160 Myr of simulation time and

scaling those numbers to account both for YORP cycles (for example, for $D > 1$ km objects, we reduced the population by a factor of 3) and the collisional evolution results described in Fig. 3 (for example, for $D > 1$ km objects, we reduced the population by $\sim 40\%$), we estimate that the current representation of $D > 1$ km BAF members in the NEO population is $\sim 0.18\%$ that of the initial population. Hence, if the BAF initially had $(1.36 \pm 0.3) \times 10^5$ objects of this size (Fig. 2), the current number of km-sized BAF objects in the NEO population today should be 240 ± 50 . For comparison, there are $\sim 1,100$ km-sized NEOs with $a < 7.4$ AU, with half being dark C- or X-type asteroids or dormant Jupiter-family comets^{11,34–36}. This implies that of the NEOs with $D > 1$ km, the BAF may be responsible for 40% of all C- and X-types and 20% of the entire population.

Impact flux on terrestrial planets

To compute the impact rate of BAF members on the terrestrial planets, we numerically integrated 9,024 test bodies placed inside the J7:2/M5:9. Our runs again used SWIFT-RMVS3 (ref. 29) but did not include the Yarkovsky effect because of its limited importance for planet-crossing asteroids. Tracking their evolution throughout the inner Solar System, we found that 1.7% hit Earth over 200 Myr of evolution (see Supplementary Discussion).

Figure 4 shows the expected impact rate distribution of BAF members on Venus, Earth and Mars for $D > 1$ km, $D > 5$ km and $D > 10$ km asteroids. A Monte Carlo code was used to combine the number of BAF objects at time t after the family-forming event with two probability distributions, one describing the likelihood that BAF objects will become trapped in the J7:2/M5:9 at time t and the other describing the likelihood that objects in the J7:2/M5:9 will hit a planet at time t . For Earth impacts, the BAF asteroid shower peaked at $t \approx 40$ Myr and then slowly decayed to the present day. This peak is consistent with an inferred change in the terrestrial impactor flux that occurred 100–120 Myr ago (see, for example, ref. 4). At this time, the number of km-sized BAF asteroids in NEO space was ~ 3 times its current value; this implies that the total NEO population was 1.4 times its present-day size, with $\sim 65\%$ being C- or X-type asteroids.

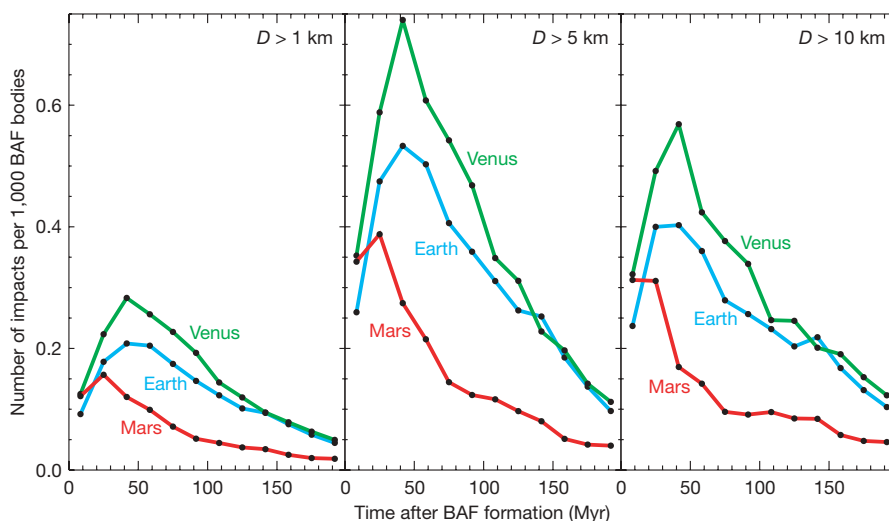


Figure 4 | The impact rate of Baptista fragments on Venus, Earth and Mars. For reference, the ordinate values were normalized to be the number of impacts per 1,000 BAF objects, whereas the abscissa is time after the Baptista family-formation event. This plot was created by combining results from two different numerical simulations within a Monte Carlo code (see Supplementary Discussion). In the first simulation, we created a model BAF and tracked the time needed for test asteroids with $D = 1, 5$ and 10 km to become trapped in the J7:2/M5:9. Scaling for the YORP effect, we found that 13%, 23% and 19%, did this, respectively, over 200 Myr of evolution. The $D > 1$ km value was then further scaled to account for the depletion of these objects by collisional evolution as a function of time (Fig. 3). For the second simulation, we numerically integrated test bodies placed inside the

J7:2/M5:9 and determined both the fraction that struck Venus, Earth and Mars, and the times of these impacts. These runs included perturbations from the planets Venus through to Neptune but excluded the Yarkovsky effect. We found that 2.1%, 1.7% and 0.9% hit Venus, Earth and Mars over 200 Myr of evolution. Using our Monte Carlo code, we randomly added times together from both simulations and then tabulated the results over 500,000 simulations. Assuming that the BAF is 160 Myr old, the number of $D > 1$ km, $D > 5$ km and $D > 10$ km impacts on Earth over this time are 200 ± 60 , 6 ± 2 and 1 ± 1 , respectively, whereas those from the background are 260 ± 20 , 3 ± 2 and 0.5 ± 0.7 , respectively. Hence, the Baptista asteroid shower increased the impact flux on the Earth and Moon by a factor of 2–3 for multi-kilometre projectiles.

When this information is combined with the BAF escape rates provided above and the evolved SFD shown in Fig. 3, we get impact rates that indicate the Earth should have been hit over the past 160 Myr by 200 ± 60 objects with $D > 1$ km, 6 ± 2 objects with $D > 5$ km, and 1 ± 1 objects with $D > 10$ km (see Supplementary Discussion). For the latter value, assuming Poisson statistics, the formal probability that one or more $D > 10$ km objects has hit the Earth over the past 160 Myr is 60%. The relationship between the largest BAF impactor and the K/T event is discussed below.

Numerical calculations indicate that $D > 1$ km asteroids currently strike the Earth, on average, every 0.5 Myr (refs 11, 35, 36). If we remove the estimated BAF contribution, this value changes to ≥ 0.63 Myr. Inserting these rates into a Monte Carlo code, we find that over the past 160 Myr, km-sized BAF members have struck the Earth 200 ± 60 times whereas those from the background population have hit 260 ± 20 times (see also Fig. 4 legend). These values are in excellent agreement with the approximate factor of 2 increase in the km-sized impactor flux inferred from terrestrial and lunar craters and the lunar glass spherule data^{3–6,8,9}. It also suggests that the majority of craters formed ~ 100 – 120 Myr ago were made by C- and X-type projectiles, a prediction that should be testable.

Origin of the K/T impactor

The K/T mass extinction event, which was by far the largest of the Mesozoic and Cenozoic eras (that is, the last 250 Myr), has been strongly linked to the formation of the 180-km-wide Chicxulub crater that formed 65 Myr ago^{37,38}. It was produced by the impact of a ~ 10 -km projectile (see refs 37 and 39, also <http://www.lpl.arizona.edu/tekton/crater.html>). The two possible sources for this event are the BAF asteroid shower and a background NEO population that includes comets. Here we estimate the relative likelihood that either produced the Chicxulub crater.

First it is useful to describe what we know of the Chicxulub impactor. Chromium found in sediment samples taken from different K/T boundary sites^{40,41}, as well as a meteorite found in K/T boundary sediments from the North Pacific Ocean⁴², suggest the impactor was a CM2-type carbonaceous chondrite. This classification is consistent with the C- and X-type taxonomy of the BAF and rules out the possibility that the K/T impactor came from an S- or V-type asteroid⁴³. Hence, if the K/T impactor came from the background NEO population, it had to come out of the relatively small fraction that had C-type, X-type or comet taxonomy 65 Myr ago. As described above, this would be equivalent to $50\% - 20\% \approx 30\%$ of the present-day population. As a check on this value, we examined all of the $D \geq 10$ km NEOs with known taxonomic types; according to our model, nearly all should now be background objects. We found 6 of 7 (86%) have S-type taxonomy, consistent with our estimate that $>70\%$ should be this way.

Next, we examined the impact flux from each source. The terrestrial impact rate of $D > 10$ km projectiles from the BAF is ~ 1 over the past 160 Myr. The frequency of comparable impacts from the background NEO population is more difficult to calculate for several reasons: (1) the current orbital and size distribution of large NEOs is not necessarily representative of time-averaged conditions^{2–6}, (2) no $D > 10$ km impactors are currently on Earth-crossing orbits¹⁸, and (3) the population of large terrestrial craters is incomplete and suffers from selection effects (for example, removal of craters by erosion, and observing biases)^{3,4}.

To avoid these problems, we computed the current escape rate of $D > 10$ km bodies from the inner main belt and their estimated terrestrial impact rate. This source is thought to dominate the large body impactor flux from the background NEO population with $a < 7.4$ AU, so the impact rates from several other background sources (for example, outer main belt, active and dormant Jupiter-family comets) can be folded into this value^{2–11}. According to our model results, the BAF contribution to the $D > 10$ km flux is minimal at present, such that the large object flux is arguably more likely to represent

time-averaged conditions over the past several hundred Myr than existing NEOs. Numerical results² indicate that one $D > 10$ km body escapes the inner main belt every ~ 4 Myr and that 1.25% of them strike the Earth⁴⁴. This translates into an approximate impact rate of one per 350 Myr. If we then assume that $<30\%$ of the background were C- and X-type asteroids or Jupiter family comets 65 Myr ago, and only 40% of those objects had CM-like compositions, estimates that are consistent with present-day meteorite fall statistics⁴⁵, the impact flux drops to one per 3,500 Myr. A comparable value was found when we examined the possibility that the K/T impactor came from nearly isotropic comets (see Supplementary Discussion). Combining these rates, we predict that primitive K/T-like projectiles from the background population strike the Earth once per 1,800–2,600 Myr.

Using these impact rates as input for a Monte Carlo code, we find there is a $\lesssim 10\%$ chance that the K/T impactor was derived from the background and a $\gtrsim 90\%$ chance it came from the BAF. Accordingly, we predict that the most likely cause of the K/T mass extinction event was a collision between the Earth and a large fragment from the Baptistina asteroid shower.

Support for a connection between CM meteorites and the BAF can be found by examining what is known about the meteorite delivery process. Collision and dynamical modelling work on asteroid families tell us that large families in the innermost region of the main belt, which produce numerous meteoroids via a collisional cascade and have access to the most efficient dynamical pathways to reach Earth, are expected to dominate our meteorite fall statistics^{26,44}. Considering that the BAF is the only prominent C- or X-type family in this region, we predict it should be the source of the largest class of carbonaceous chondrite falls, which happens to be that of the CM meteorites⁴⁵. Accordingly, the CM meteorite and $D \approx 170$ km Baptistina parent bodies are likely to be one and the same.

Additional implications

The lunar crater Tycho is something of a statistical anomaly when one considers its large size (85 km) and relatively young age (109 ± 4 Myr; see ref. 46). Calculations suggest that Tycho was created by the impact of a $D > 4$ km projectile³⁹. Using our work, we estimate that 0.6 BAF members of this size have hit the Moon over the past 160 Myr, whereas the impact rate from the background population is one per 570 Myr (ref. 2). The contribution from nearly isotropic comets is negligible. Because there is no compositional information on the Tycho projectile, we cannot pare these numbers further. Incorporating them into a Monte Carlo code, we find there is a $\sim 70\%$ chance that the projectile that created Tycho was from the BAF rather than the background population. The fact that Tycho formed near the peak of the Baptistina asteroid shower is also suggestive of a link between the two (Fig. 4).

Similar techniques were used to investigate the formation history of the largest craters on Venus: Mead (269 km), Isabella (176 km), Meitner (151 km) and Klenova (142 km)⁴⁷. Numerical simulations show that these craters formed over the past 730 Myr (ref. 48) from roughly K/T-sized projectiles striking the venusian surface³⁹. Estimates of the background impactor flux indicate that projectiles this size strike Venus and Earth at nearly the same rate (one per ~ 350 Myr; ref. 2). Using these data as well as information from Figs 2 and 4, we found an 80% probability that at least one of these craters was produced by the Baptistina asteroid shower (see Supplementary Discussion).

Probing the martian impactor flux, we estimate that approximately 110 $D > 1$ km BAF fragments have struck Mars over the past 160 Myr. For the background population over the same size range, the interval between impacts is roughly 1.3 Myr (ref. 2). This means that the impact rate of $D > 1$ km bodies on Mars has nearly doubled over the past 160 Myr, with most impactors striking within a few tens of Myr after the Baptistina parent body disrupted (Fig. 4). It will be interesting to see how further studies of the Baptistina asteroid shower affect our understanding of recent martian chronology.

The BAF also has some surprising implications for asteroids in the main belt itself. (951) Gaspra, a $19 \times 12 \times 11$ km S-type asteroid in the Flora family, has a fresh, non-saturated population of $D < 0.6$ km craters whose SFD is far steeper ($q = -3.3 \pm 0.3$; ref. 27) than expectations based on collisional models of the main belt production population (for example, $q \approx -2.5$; ref. 1). We find that the majority of these craters may have been produced by an intense barrage of sub-100-m fragments tied to the Baptistina family-forming event²⁷. If true, the signature of this bombardment should also exist on other asteroids that once had, or still have, Baptistina-like orbits.

The interval between $D > 100$ km disruption events across the main belt is short enough (~ 200 Myr; refs 1, 24) that other asteroid showers should have occurred in the past¹⁵. The Flora family, which was created 470 Myr ago by the disruption of a $D > 200$ km S-type asteroid residing near $a \approx 2.2$ AU (refs 17, 24, 49), seems to be the best candidate to have produced a relatively recent Baptistina-like asteroid shower. The family's age has been precisely determined by linking its ejecta to the existence of tiny L-chondrite meteorites found within the limestone walls of a Swedish quarry^{16,17}. Preliminary work suggests that the Flora family lost a significant fraction of its initial population of $D > 1$ km asteroids to Mars-crossing orbits (see, for example, ref. 49). If so, this putative shower may explain the abundance of ~ 200 – 500 -Myr-old impact-derived glass spherules found in lunar soils⁹ as well as the increased number of lunar craters formed over these same ages⁸.

Received 11 April; accepted 22 June 2007.

1. Bottke, W. F. *et al.* Linking the collisional history of the main asteroid belt to its dynamical excitation and depletion. *Icarus* **179**, 63–94 (2005).
2. Bottke, W. F. *et al.* Debiased orbital and absolute magnitude distribution of the near-Earth objects. *Icarus* **156**, 399–433 (2002).
3. Shoemaker, E. M. Impact cratering through geologic time. *J. R. Astron. Soc. Can.* **92**, 297–309 (1998).
4. Grieve, R. A. F. & Shoemaker, E. M. in *Hazards Due to Comets and Asteroids* (eds Gehrels, T. & Matthews, M. S.) 417–462 (Univ. Arizona Press, Tucson, 1994).
5. McEwen, A. S., Moore, J. M. & Shoemaker, E. M. The Phanerozoic impact cratering rate: Evidence from the farside of the Moon. *J. Geophys. Res.* **102**, 9231–9242 (1997).
6. Grier, J. A. *et al.* Optical maturity of ejecta from large rayed lunar craters. *J. Geophys. Res.* **106**, 32847–32862 (2001).
7. Ward, S. N. & Day, S. Terrestrial crater counts: Evidence for a four-fold increase in bolide flux at 125 Ma. ([http://es.usc.edu/~ward/papers/crater-counts\(v1.8\).pdf](http://es.usc.edu/~ward/papers/crater-counts(v1.8).pdf)) (2007).
8. Baldwin, R. B. Relative and absolute ages of individual craters and the rate of infalls on the moon in the post-Imbrium period. *Icarus* **61**, 63–91 (1985).
9. Levine, J., Becker, T. A., Muller, R. A. & Renne, P. R. ⁴⁰Ar/³⁹Ar dating of Apollo 12 impact spherules. *Geophys. Res. Lett.* **32**, L15201, doi:10.1029/2005GL022874 (2005).
10. Dones, L. *et al.* in *Comets II* (eds Festou, M. C. *et al.*) 153–174 (Univ. Arizona Press, Tucson, 2004).
11. Stokes, G. *et al.* A study to determine the feasibility of extending the search for near Earth objects to smaller limiting magnitudes. (neo.jpl.nasa.gov/neo/report.html) (2003).
12. Levison, H. *et al.* The mass disruption of Oort cloud comets. *Science* **296**, 2212–2215 (2002).
13. Tagle, R. & Claeys, P. Comet or asteroid shower in the Late Eocene? *Science* **305**, 492 (2004).
14. Maier, W. D. *et al.* Discovery of a 25-cm asteroid clast in the giant Morokweng impact crater, South Africa. *Nature* **441**, 202–206 (2006).
15. Zappalà, V., Cellino, A., Gladman, B. J., Manley, S. & Migliorini, F. Asteroid showers on Earth after family breakup events. *Icarus* **134**, 176–179 (1998).
16. Schmitz, B., Haggström, T. & Tassinari, M. Sediment-dispersed extraterrestrial chromite traces a major asteroid disruption event. *Science* **300**, 961–964 (2003).
17. Nesvorný, D., Vokrouhlický, D., Bottke, W. F., Gladman, B. J. & Haggström, T. Express delivery of fossil meteorites from the inner asteroid belt to Sweden. *Icarus* **188**, 400–413 (2007).
18. Knežević, Z. & Milani, A. Proper element catalogs and asteroid families. *Astron. Astrophys.* **403**, 1165–1173 (2003).
19. Mothé-Diniz, T., Roig, F. & Carvano, J. M. Reanalysis of asteroid families structure through visible spectroscopy. *Icarus* **174**, 54–80 (2005).
20. Morbidelli, A. & Nesvorný, D. Numerous weak resonances drive asteroids toward terrestrial planets orbits. *Icarus* **139**, 295–308 (1999).
21. Bottke, W. F., Vokrouhlický, D., Rubincam, D. P. & Nesvorný, D. The Yarkovsky and YORP effects: Implications for asteroid dynamics. *Annu. Rev. Earth Planet. Sci.* **34**, 157–191 (2006).
22. Vokrouhlický, D., Brož, M., Bottke, W. F., Nesvorný, D. & Morbidelli, A. Yarkovsky/YORP chronology of asteroid families. *Icarus* **182**, 111–142 (2006).
23. Ivezić, Ž. *et al.* Color confirmation of asteroid families. *Astron. J.* **124**, 2943–2948 (2002).
24. Durda, D. D. *et al.* Size–frequency distributions of fragments from SPH/N-body simulations of asteroid impacts: Comparison with observed asteroid families. *Icarus* **186**, 498–516 (2007).
25. Durda, D. D. *et al.* The formation of asteroid satellites in catastrophic impacts: Results from numerical simulations. *Icarus* **167**, 382–396 (2004).
26. Bottke, W. F. *et al.* in *Dynamics of Populations of Planetary Systems* (eds Knežević, Z. & Milani, A.) 357–376 (Cambridge Univ. Press, Cambridge, UK, 2005).
27. Bottke, W. F., Vokrouhlický, D., Chapman, C. R. & Nesvorný, D. Gaspra's steep crater population was produced by a large recent breakup in the main asteroid belt. *Lunar Planet. Sci. Conf. abstr.* 2165 (2007).
28. Britt, D. T. *et al.* in *Asteroids III* (eds Bottke, W. F. *et al.*) 485–500 (Univ. Arizona Press, Tucson, 2002).
29. Levison, H. F. & Duncan, M. J. The long-term dynamical behavior of short-period comets. *Icarus* **108**, 18–36 (1994).
30. Brož, M. *Yarkovsky Effect and the Dynamics of the Solar System*. PhD thesis, Charles Univ. (2006); (<http://sirrah.troja.mff.cuni.cz/~mira/mp/>) (2006).
31. Pravec, P., Harris, A. W. & Michalowski, T. in *Asteroids III* (eds Bottke, W. F. *et al.*) 113–122 (Univ. Arizona Press, Tucson, 2002).
32. Delbò, M. *et al.* Thermal inertia of near-Earth asteroids and implications for the magnitude of the Yarkovsky effect. *Icarus*. (in the press).
33. Vokrouhlický, D. & Čapek, D. YORP-induced long-term evolution of the spin state of small asteroids and meteoroids. Rubincam's approximation. *Icarus* **159**, 449–467 (2002).
34. Stuart, J. S. & Binzel, R. P. Bias-corrected population, size distribution, and impact hazard for the near-Earth objects. *Icarus* **170**, 295–311 (2004).
35. Morbidelli, A. *et al.* From magnitudes to diameters: The albedo distribution of near Earth objects and the Earth collision hazard. *Icarus* **158**, 329–342 (2002).
36. Binzel, R. P. *et al.* Observed spectral properties of near-Earth objects: results for population distribution, source regions, and space weathering processes. *Icarus* **170**, 259–294 (2004).
37. Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. V. Extraterrestrial cause for the Cretaceous Tertiary extinction. *Science* **208**, 1095–1108 (1980).
38. Kring, D. A. The Chicxulub impact event and its environmental consequences at the K/T boundary. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* (in the press).
39. Melosh, H. J. *Impact Cratering: A Geologic Process* (Oxford Univ. Press, New York, 1989).
40. Shukolyukov, A. & Lugmair, G. W. Isotopic evidence for the Cretaceous-Tertiary impactor and its type. *Science* **282**, 927–930 (1998).
41. Trinquier, A., Birck, J.-L. & Allègre, J. C. The nature of the K/T impactor. A ⁵⁴Cr reappraisal. *Earth Planet. Sci. Lett.* **241**, 780–788 (2006).
42. Kyte, F. T. A meteorite from the Cretaceous-Tertiary boundary. *Nature* **396**, 237–239 (1998).
43. Clark, B. E. in *Asteroids III* (eds Bottke, W. F. *et al.*) 585–599 (Univ. Arizona Press, Tucson, 2002).
44. Bottke, W. F. *et al.* Iron meteorites as remnants of planetesimals formed in the terrestrial planet region. *Nature* **439**, 821–824 (2006).
45. Burbine, T. H. *et al.* in *Asteroids III* (eds Bottke, W. F. *et al.*) 653–667 (Univ. Arizona Press, Tucson, 2002).
46. Stöffler, D. & Ryder, G. Stratigraphy and isotope ages of lunar geologic units: Chronological standard for the inner solar system. *Space Sci. Res.* **96**, 9–54 (2001).
47. Herrick, R. R. *et al.* in *Venus II* (eds Bougher, S. W. *et al.*) 1015–1046 (Univ. Arizona Press, Tucson, 1997).
48. Korycansky, D. G. & Zahnle, K. J. Modeling crater populations on Venus and Titan. *Planet. Space Sci.* **53**, 695–710 (2005).
49. Nesvorný, D. *et al.* The Flora family: A case of the dynamically dispersed collisional swarm? *Icarus* **157**, 155–172 (2002).
50. Bottke, W. F. *et al.* Velocity distribution among colliding asteroids. *Icarus* **107**, 255–268 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Beatty, C. Chapman, L. Dones, D. Durda, B. Enke, D. Kring, F. Kyte, M. Gounelle, R. Grimm, A. Harris, H. Levison, A. Morbidelli, A. Rubin, E. Scott, A. Stern and M. Zolensky for discussions and comments. We also thank G. Williams of the Minor Planet Center for computing revised H values and observational errors for the Baptistina family. The work of W.F.B. and D.N. on this project was supported by NASA's Origins of Solar System, Planetary Geology and Geophysics, and Near-Earth Objects Observations programmes. D.V. was supported by the Grant Agency of the Czech Republic.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.F.B. (bottke@boulder.swri.edu).

ARTICLES

Natural history and evolutionary principles of gene duplication in fungi

Ilan Wapinski^{1,2,3}, Avi Pfeffer³, Nir Friedman⁴ & Aviv Regev^{1,5}

Gene duplication and loss is a powerful source of functional innovation. However, the general principles that govern this process are still largely unknown. With the growing number of sequenced genomes, it is now possible to examine these events in a comprehensive and unbiased manner. Here, we develop a procedure that resolves the evolutionary history of all genes in a large group of species. We apply our procedure to seventeen fungal genomes to create a genome-wide catalogue of gene trees that determine precise orthology and paralogy relations across these species. We show that gene duplication and loss is highly constrained by the functional properties and interacting partners of genes. In particular, stress-related genes exhibit many duplications and losses, whereas growth-related genes show selection against such changes. Whole-genome duplication circumvents this constraint and relaxes the dichotomy, resulting in an expanded functional scope of gene duplication. By characterizing the functional fate of duplicate genes we show that duplicated genes rarely diverge with respect to biochemical function, but typically diverge with respect to regulatory control. Surprisingly, paralogous modules of genes rarely arise, even after whole-genome duplication. Rather, gene duplication may drive the modularization of functional networks through specialization, thereby disentangling cellular systems.

Gene duplication and loss are major forces of evolutionary innovation, facilitating the development of new functions and pruning of old ones^{1,2}. Nonetheless, the natural history of gene duplication and loss is poorly understood. What classes of genes readily evolve through duplication and loss? Do whole-genome duplication events reshape the genome in a qualitatively distinct way? What innovations typically arise from gene duplication events? Studies addressing such questions^{3–11} have been limited by the difficulty of tracing the exact evolutionary history of genes.

The growing availability of sequenced genomes enables the direct reconstruction of a genome-wide history of gene duplication and loss across species^{3,7}. Here, we describe a computational method for reconstructing this history and apply it to the genomes of seventeen Ascomycota fungi spanning 300 million years of evolution^{12–21}. The results suggest evolutionary principles applicable for fungi and possibly more generally.

Method for identifying orthologues and paralogues

Systematic study of gene duplication and loss requires reliable resolution of gene orthology and paralogy, a notoriously difficult problem^{22–31}. We designed SYNERGY, a scalable method for resolving gene ancestry for all genes across multiple genomes (Fig. 1, Supplementary Fig. 1)³². The input is a species phylogeny and, for each extant species, the sequences of predicted genes and their chromosomal positions. SYNERGY partitions these genes into ‘orthogroups’. Each orthogroup consists of all (and only) the genes descended from a single ancestral gene in their last common ancestral species, and is associated with a gene tree that describes the history of speciation, duplication and loss events for its genes (Methods).

An orthogroup catalogue for Ascomycota fungi

We applied SYNERGY to the complete set of 121,050 predicted protein-coding genes from seventeen genomes of Ascomycota fungi,

including the model organisms *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (Fig. 2a, Methods). The phylogeny includes a whole-genome duplication (WGD) event^{14,19,33} (Fig. 2a, red star). SYNERGY produced a catalogue of 30,110 orthogroups (Fig. 2b). Of these, 19,006 were singleton genes with no recognizable orthologues (Supplementary Note 1). We further analysed the 11,103 multigene orthogroups. The orthogroups and trees are available at <http://www.broad.mit.edu/regev/orthogroups/>.

SYNERGY made high-quality predictions by several benchmarks (Methods, Supplementary Notes 2 and 3). To test sensitivity to the input quality, we applied SYNERGY to different subsets of organisms and of genes in each genome. We examined how each orthogroup was reconstructed under these perturbations, deriving four confidence measures for each orthogroup. Overall, SYNERGY was remarkably robust (Supplementary Note 2). SYNERGY’s predictions also agree well with those of two independent manual assignments of orthology and paralogy^{21,33} (Supplementary Note 3). Finally, SYNERGY showed high specificity and sensitivity on data attained by forward simulated evolution.

Gene duplication and loss across Ascomycota evolution

The reconstructed orthogroups show a range of evolutionary patterns. These are summarized by the extended phylogenetic profile (EPP) of each orthogroup, defined as the number of genes present in each extant and ancestral species. For example, ‘uniform’ orthogroups (Fig. 2c), with no duplication or loss events, have EPPs consisting only of ones. Other orthogroups exhibit duplications (Fig. 2d, red star) or losses (Fig. 2d, blue strikes) and their EPPs may consist of noughts, ones, twos, and so on (Fig. 2e). An orthogroup with at least one gene present in all species (Supplementary Fig. 2b) is ‘persistent’. From the EPP, we can derive an extended copy-number variation profile that records the change in copy number at each position in the species tree (Supplementary Fig. 2).

¹Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²FAS Center for Systems Biology, Harvard University, 7 Divinity Avenue, ³School of Engineering and Applied Sciences, Harvard University, 33 Oxford Street, Cambridge, Massachusetts 02138, USA. ⁴School of Computer Science and Engineering, Hebrew University, Jerusalem 91904, Israel. ⁵Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

By tallying these profiles from all orthogroups, we find the numbers of genes, appearances, duplications and losses that occurred throughout Ascomycota evolution (Fig. 2f, Supplementary Fig. 3a). The 5,972 (54%) orthogroups present in the clade spanning the Hemiascomycota and Euascomycota were defined as ‘ancestral’, accounting for 6,047 genes in the reconstructed last common ancestor. 4,873 (84%) of *Saccharomyces cerevisiae* genes belong to these ancestral orthogroups. The ancestral orthogroups are enriched in essential *S. cerevisiae* genes; 1,008 of 1,047 genes essential for growth in rich medium are ancestral ($P < 10^{-30}$, Fischer’s exact test) as are 668 of 730 genes essential only for growth in other con-

ditions³⁴ ($P < 10^{-5}$). Nevertheless, 36 essential genes are not ancestral (for example, 9/80 spindle pole body proteins, $P < 10^{-6}$), suggesting that new essential functions can arise, albeit rarely (Supplementary Note 4).

Orthogroups that are not ancestral ‘appear’ at specific points in the phylogeny and reveal evolutionary innovations. For example, the clade spanning *S. cerevisiae* and *Kluyveromyces waltii* is marked by appearing orthogroups with *S. cerevisiae* genes related to meiosis and sporulation (51/166 sporulation genes, $P < 10^{-6}$), including the master meiosis regulator IME1³⁵ (Supplementary Fig. 2a). The Euascomycota clade contains many appearances (3,726 orthogroups, 72% of all appearing orthogroups); roughly half show no similarity to other orthogroups or to a more distant fungus, *Cryptococcus neoformans* (Supplementary Fig. 3d), demonstrating extensive genomic innovation within the Euascomycota¹⁶.

We also find coordinated gene losses, indicating major changes in biological processes. For example, *Yarrowia lipolytica* has significant losses in orthogroups containing meiotic recombination genes ($P < 10^{-5}$). Interestingly, the genes lost in *Candida glabrata* extensively overlap those independently lost in the ancestor of *Candida albicans* and *Debaryomyces hansenii* ($P < 10^{-20}$), possibly reflecting the fact that these are all opportunistic or occasional human pathogens.

Copy number volatility corresponds to a functional dichotomy

The observed variation in copy number changes among orthogroups is inconsistent with random expectation (Fig. 3a, Methods). We assigned a volatility score to each orthogroup depending on the number and phylogenetic position of duplication and loss events, with 1,018 uniform orthogroups at one end of the scale and 313 ‘volatile’ orthogroups (score > 3 s.d. above the mean) at the other (Fig. 3a). Evolutionary forces have acted very differently on these two classes: the uniform and volatile orthogroups show diametrically opposed patterns in their function, regulation and essentiality in *S. cerevisiae* (Fig. 3, Supplementary Table 1).

We first tested for functional distinctions between uniform and volatile orthogroups, based on gene ontology annotations in *S. cerevisiae*³⁶. Volatile orthogroups are enriched ($P < 10^{-5}$) for genes that encode peripheral transporters, receptors and cell wall proteins, and genes that participate in stress responses. In contrast, uniform orthogroups are enriched ($P < 10^{-5}$) for genes involved in essential growth processes, genes residing in the nucleus, nucleolus, mitochondrion, endoplasmic reticulum and Golgi apparatus, and genes essential for viability.

We next examined whether the evolutionary dichotomy is also aligned with the transcriptional program of *S. cerevisiae*. Using data from 1,216 gene expression profiles, we organized *S. cerevisiae* genes into a hierarchy of 163 transcriptional modules, each containing functionally related genes with a coherent expression pattern³⁷ (Fig. 4a, Supplementary Fig. 4, Supplementary Table 2, and

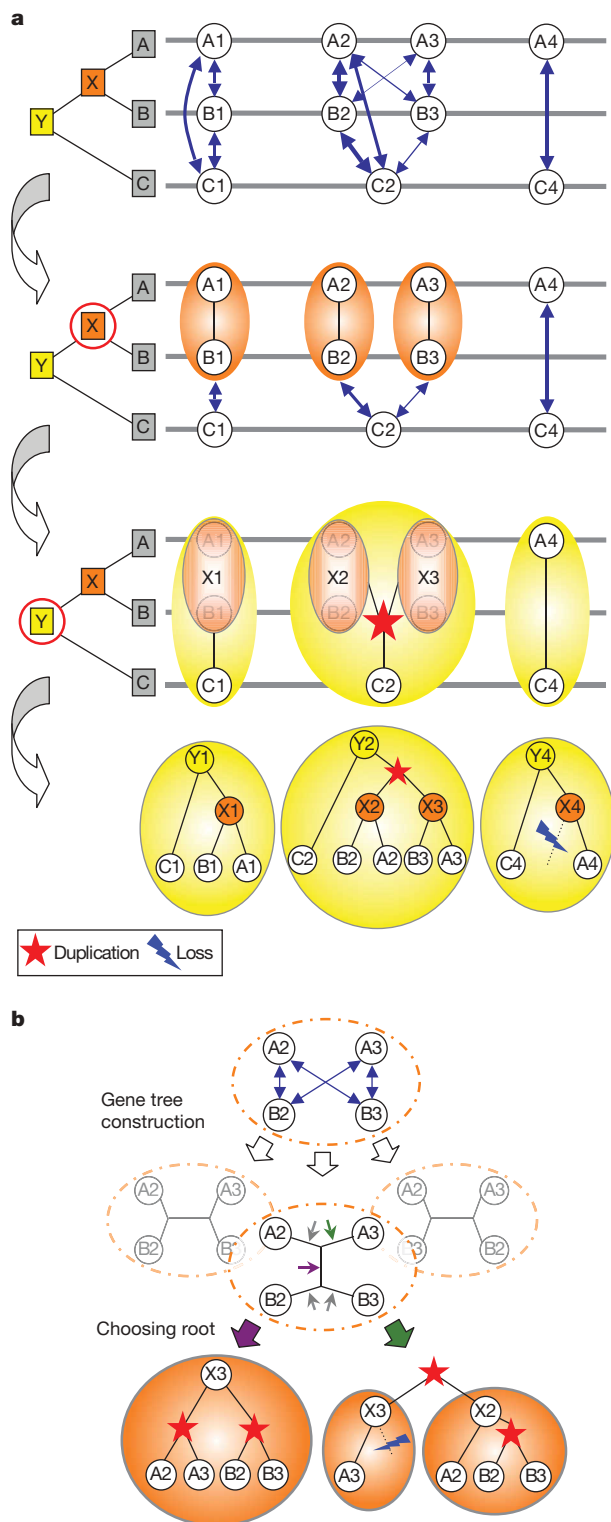


Figure 1 | The SYNERGY algorithm. **a**, Orthogroup construction. SYNERGY starts (top) with a collection of genes (A1, B1, C1 and so on), their chromosomal order (grey lines) and sequence distances (blue arrows; arrows of the same thickness have similar sequence distances). It then builds orthogroups as it climbs the species tree. First, it collects the genes in species A and B that share a common ancestor in species X (second panel, orange ovals). Then, it merges orthogroups formed in the previous stage with the genes in C, resulting in new orthogroups representing ancestral genes in species Y (third panel, yellow ovals). The orthogroups assembled at each stage are associated with gene trees reflecting divergence, duplication and loss events (bottom). **b**, Gene tree reconstruction and refining orthogroup assignments. An unrooted phylogeny is reconstructed for the genes and sub-orthogroups in each putative orthogroup (dashed oval). Some rootings (purple arrow) indicate that all the genes descended from a common ancestor (for example, X3, bottom left). Others (green arrow) show that a duplication occurred at the root of the gene tree (for example, X2 and X3, bottom right). In the latter case, the orthogroup is partitioned before proceeding.

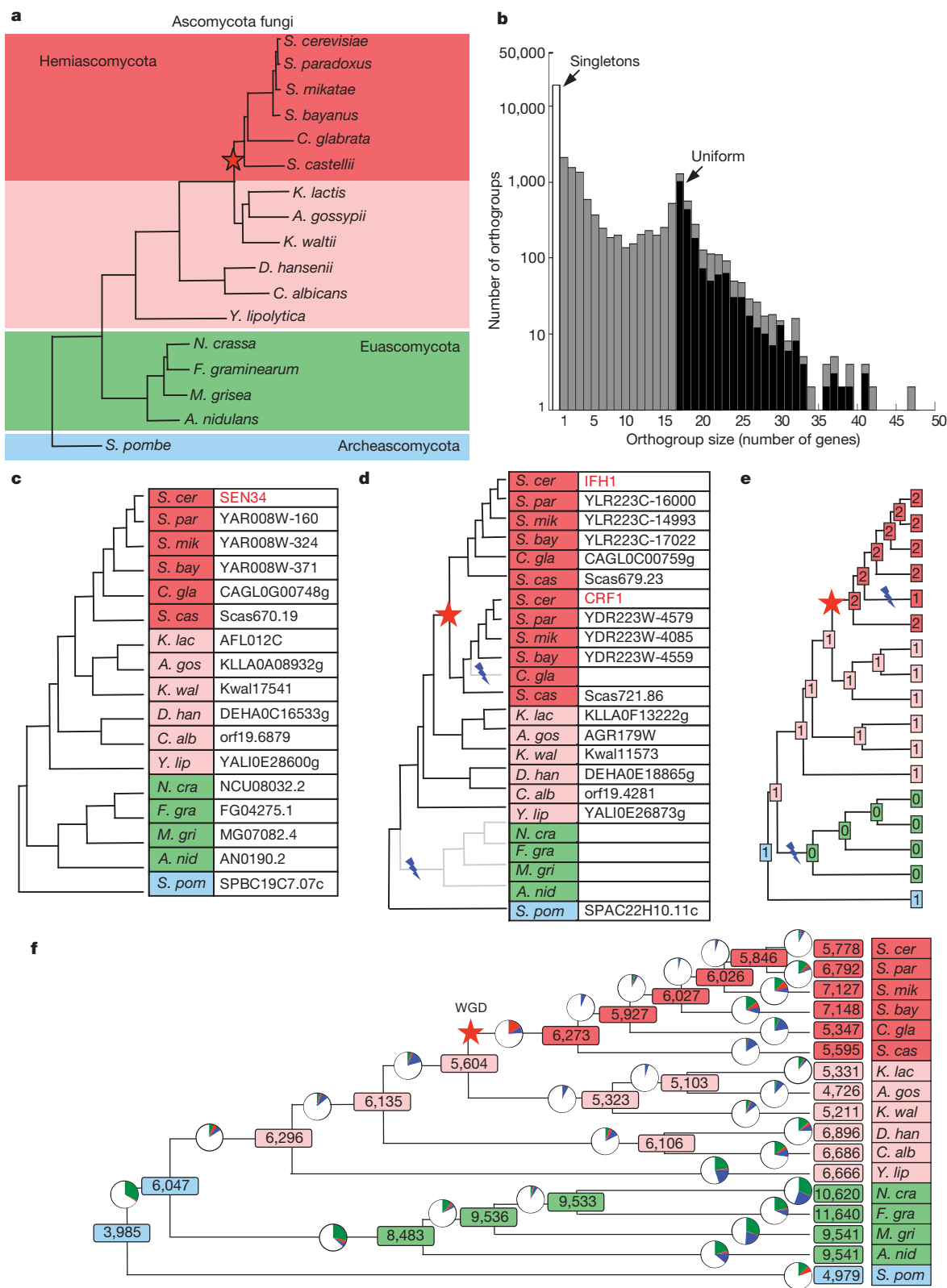


Figure 2 | A gene ancestry catalogue for Ascomycota fungi. **a**, Species tree showing the Hemiascomycota (pink), Euscomycota (green), and Archeascomycota (blue) clades, the WGD (red star), and post-WGD species (darker pink). **b**, Distribution of orthogroup sizes. Grey bars, total number of orthogroups of a certain size (number of genes). Black bars, the number of persistent orthogroups of a certain size. The uniform arrow points to orthogroups, which are persistent orthogroups with exactly 17 genes. The singleton arrow points to orthogroups with a single gene. **c**, A uniform orthogroup. The topology of the gene tree (left panel) is identical to that of

the species tree. **d**, A non-uniform orthogroup with a single duplication event (red star) and two loss events (blue strikes). **e**, The extended phylogenetic profile of the orthogroup in **d** summarizes the number of genes at each extant and ancestral species in the tree (numbered boxes). **f**, Reconstructed evolutionary events and gene counts. Each species is annotated with the number of known or reconstructed genes (rectangles). Pie-charts on branches denote the fraction of appearing (green), duplicated (red), and lost (blue) genes inferred in the corresponding branch (see Supplementary Fig. 3 for exact numbers).

Methods). Analysing the volatility scores of these transcriptional modules, we see that the evolutionary dichotomy follows the regulatory branches (Fig. 4c, d, Supplementary Table 3). Modules in the 'cell cycle and meiosis' and 'fundamental processes' branches are enriched ($P < 10^{-4}$) for uniform and persistent orthogroups, whereas modules in the 'development' and 'stress and carbohydrate metabolism' branches are enriched ($P < 10^{-4}$) for volatile orthogroups.

These distinctions indicate a limit not only on the ability to lose key growth genes, but also on the ability to maintain them in duplicate. As suggested by the "gene balance hypothesis"⁹, one reason for this

may be that such genes often encode components of essential cellular machineries requiring stoichiometric balance. Indeed, we found that *S. cerevisiae* genes encoding core components of protein complexes³⁸ are enriched in uniform orthogroups (338/844 complex 'core' genes, $P < 10^{-32}$). Furthermore, uniform and persistent orthogroups are enriched for *S. cerevisiae* genes displaying haploinsufficiency³⁹ ($P < 10^{-4}$ and $P < 10^{-6}$, respectively). However, the dichotomy extends beyond protein complexes to many cellular processes and includes orthogroups with moderately low and high volatility scores (Fig. 3b), suggesting a general principle affecting the vast majority of genes in the genome.

To test whether differential flexibility in copy number between uniform and volatile orthogroups reflects global functional constraints, we examined the variation in their respective transcripts and proteins. We found that the volatile orthogroups are enriched in genes whose expression changes significantly in response to many single-gene knockouts⁴⁰ ($P < 10^{-5}$ – 10^{-22} ; notably deletions of chromatin modifiers), genes with noisy levels of protein abundance within isogenic *S. cerevisiae* cells⁴¹ ($P < 10^{-4}$), genes the transcription of which is regulated through the SAGA complex and the TATA box⁴² ($P < 10^{-9}$), and genes with variable RNA expression across species⁴³ ($P < 10^{-11}$) (Fig. 3, Supplementary Table 1). Conversely, the uniform orthogroups are enriched in genes whose expression is largely unchanged in response to single-gene knockouts, genes whose protein levels tend to be tightly controlled ($P < 10^{-8}$), genes whose transcription is TATA-independent and regulated through TFIID⁴² ($P < 10^{-24}$), and genes whose RNA expression shows less variation across species ($P < 10^{-15}$) (Fig. 3, Supplementary Table 1).

These results highlight a general bipolar principle that governs tolerance to duplications and losses. Copy-number variation in stress-responsive genes may not only be tolerable but beneficial, allowing adaptation to diverse ecological niches. In contrast, genes essential for cell growth, including those necessary for intricate complexes, cannot readily tolerate such noise and tend not to evolve by gradual duplication and loss. This evolutionary dichotomy aligns closely with a bipolarity in gene function, transcriptional program and expression noise across cells, strains and species^{41,43}, all reflecting similar functional constraints on the amount of gene products in the cell. Furthermore, shared functional constraints on copy-number variation also manifest in remarkably synchronized and concerted patterns of specific duplications and losses in many orthogroups harbouring functionally related or interacting genes (Supplementary Note 5, Supplementary Figs 5 and 6, and Supplementary Table 4).

Whole-genome duplication alters the nature of duplication

We next explored whether these functional principles generalize to the WGD event. We found that duplications associated with the WGD show a strikingly different pattern (Fig. 4e, Supplementary Table 3): many transcriptional modules that maintain little duplication elsewhere in the phylogeny are associated with a high level of

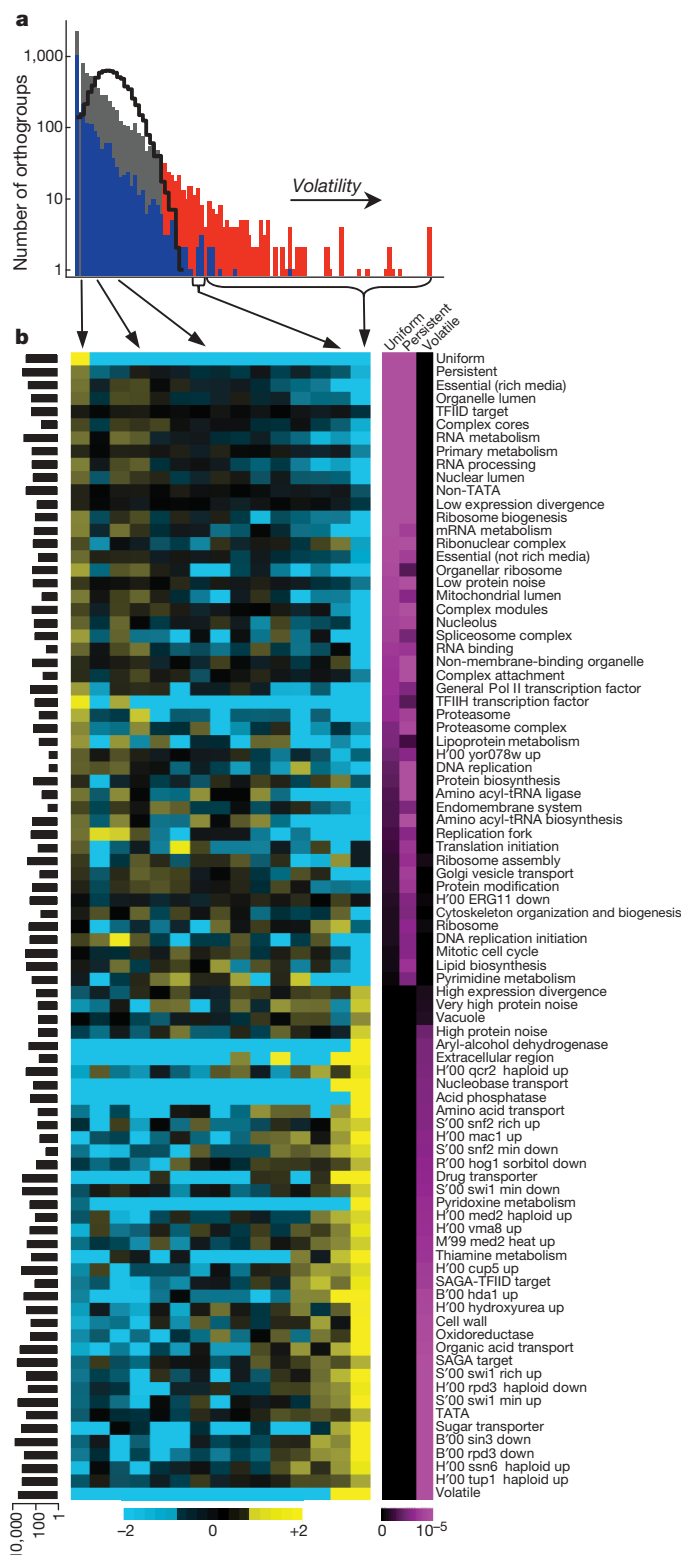


Figure 3 | A functional dichotomy of uniform, persistent and volatile orthogroups. **a**, Distribution of orthogroup volatility scores. Uniform orthogroups (leftmost blue column) are the least volatile. Orthogroups whose score is above three standard deviations from the expected mean are labelled as volatile (red columns). The remainder of the distribution is shown (grey bars) and the distribution of scores for persistent orthogroups is superimposed (blue bars). The expected distribution when sampling random orthogroups from the evolutionary model is shown as a black line. **b**, Gene class annotations that are significantly enriched among uniform, persistent or volatile orthogroups (purple, colour scale is saturated at $P < 10^{-5}$). The functional and mechanistic dichotomy between volatile and non-volatile orthogroups largely reproduces along the range of volatility scores (left columns are bins of orthogroups with similar volatility scores; rows are significant annotations). Higher (yellow) and lower (blue) relative enrichment compared to the expected enrichment in the class is shown. The colour scale is saturated at 2-fold. Class sizes are shown on the left (black bars).

volatility during the WGD. Examples include 'ribosome biogenesis' in all post-WGD lineages ($P < 10^{-5}$) and 'ER protein modification' in *Saccharomyces castellii* ($P < 10^{-3}$).

Thus, the WGD is associated with a qualitatively different pattern of duplication. The 'gene balance hypothesis'^{4,9} postulates that this effect is due to post-WGD retention of paralogues for all members of a complex. Indeed, *S. cerevisiae* genes with either haploinsufficiency³⁹ or overexpression⁴⁴ phenotypes are enriched in orthogroups that duplicated only in the WGD ($P < 10^{-9}$). Furthermore, several modules representing essential machineries are enriched for WGD paralogues⁴⁵ (for example, 'rRNA biogenesis' $P < 10^{-5}$, and 'ribosome', $P < 10^{-36}$). However, the expanded scope of the WGD is observed beyond complexes within more volatile modules. The simultaneous duplication of all genes in a module in a WGD may permit retention of paralogues in orthogroups that are otherwise constrained against duplication, and may be a principal way in which WGD events contribute to evolutionary innovations^{4,9}.

Gene duplication results in limited biochemical divergence

We next explored the types of functional innovations that arise from gene duplications. In principle, both paralogues can either 'retain'

the same function (Fig. 5a) or one (or both) can 'migrate' to assume a distinct function (Fig. 5b, c). Migration can either reflect the development of a novel function (neofunctionalization¹¹, Fig. 5b) or a division of labour, in which each paralogue assumes only some functions of the ancestral gene (subfunctionalization¹¹, Fig. 5c). Given the long-postulated importance of gene duplication in innovation^{1,2}, we hypothesized that migration would be the predominant evolutionary mode.

We quantified the extent to which paralogous gene pairs remain in or migrate from a variety of gene classes in *S. cerevisiae* (gene ontology functional categories, genes with shared regulatory mechanisms, protein complexes and transcriptional modules). We calculated the fraction of paralogous pairs that are retained within each class (Methods, and Supplementary Figs 7, 8). To avoid confounding factors, we studied only cases in which both paralogues had been annotated and the annotation had not been inferred solely from sequence similarity. Surprisingly, our analysis shows that paralogous pairs rarely migrate between functional gene ontology categories (Supplementary Figs 7–9, and Supplementary Table 5). The retention rate is highest for the 'molecular function' categories (92%) and somewhat lower for 'biological process' (85%) and 'cellular component' (81%) categories.

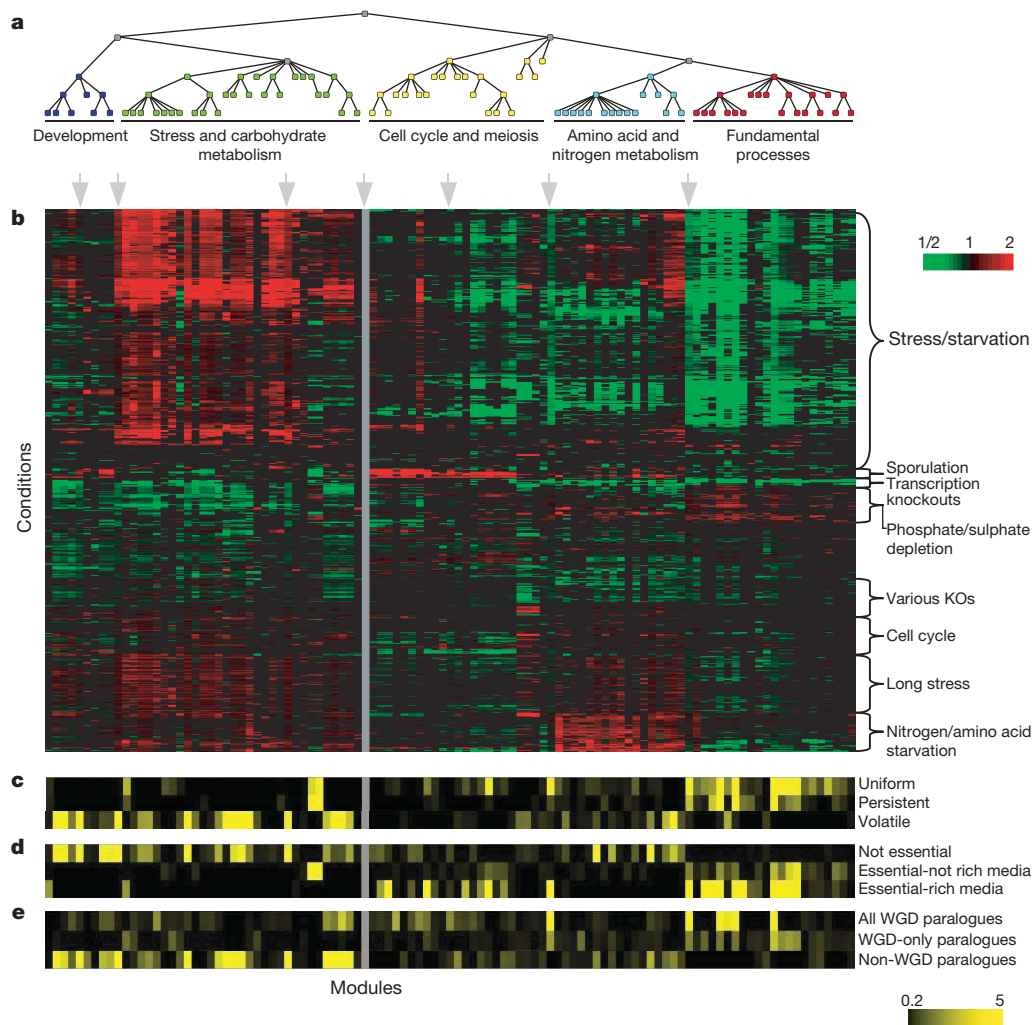


Figure 4 | Evolutionary profiles correspond to the hierarchical modular organization of the yeast transcriptional system. **a**, A functional hierarchy of *S. cerevisiae* transcriptional modules. **b**, Expression patterns of modules. Shown is the average expression of module genes (columns) in each expression array (rows, with main condition names marked on right). Red, induced; green, repressed; black, no significant change. The colour scale is saturated at 2-fold. **c–e**, Evolutionary characterization of modules. Enrichment significance (negative log P -value; yellow indicates significant;

the colour scale is saturated at 10^{-5}) for the projected orthogroup classes from each module (columns) against various phylogenetic attributes (rows). **c**, Growth modules are enriched for uniform and persistent genes, with the notable exception of amino acid metabolism modules; whereas stress modules are enriched in volatile genes. **d**, Essential genes are enriched in modules from the 'growth', but not the 'stress', groups. **e**, This dichotomy is violated in the WGD.

To reveal innovation at a finer resolution than the relatively coarse functional classes, we measured the fraction of shared interaction partners for each pair of paralogues in molecular networks. We examined both biochemical networks⁴⁶ of physically interacting proteins (reflecting molecular function), and genetic networks⁴⁶ of synthetic phenotypes in double mutants (reflecting biological processes). We find that in both networks roughly half of the paralogues share a significantly high proportion of their interaction partners (136/318 pairs in the genetic network and 225/543 in the biochemical network), much more than would be expected in comparable (degree-preserving) random networks (Methods, Supplementary Fig. 10b, f). Thus, many pairs show little migration from their pre-duplication organization (Supplementary Fig. 11a), supporting the broader result at both the biochemical and the functional level.

The remaining paralogues typically share no interaction partners and may indicate neofunctionalization (Supplementary Fig. 10a, e). Indeed, paralogous pairs had more biochemical interaction partners than would be expected by chance¹¹ (11.64 ± 24.73 versus 6.99 ± 12.66 ; $P < 10^{-3}$, Mann–Whitney *U*-test), providing global evidence for neofunctionalization. Many of these ‘disjoint paralogue pairs’ are dispersed in the biochemical network (78% are separated

by four or more proteins), implying divergence in molecular function (Supplementary Fig. 11b). In contrast, half of them are immediate neighbours in the genetic network (Supplementary Fig. 11c), suggesting a related biological process. This is consistent with the role of duplicate genes as either genetic ‘back-ups’ when their paralogues are compromised^{6,47} or with division of labour through subfunctionalization (see below).

Gene duplication innovates through regulatory divergence

Another source of innovation is regulatory divergence. We inspected the migration of paralogous pairs with respect to gene classes representing regulatory mechanisms (genes that are targets of the same transcription factor⁴⁸ or contain the same *cis*-regulatory motif⁴⁸ or RNA-binding motif⁴⁹) or expression patterns (transcriptional modules). We find that paralogous genes usually migrate with respect to these gene classes. In most cases (70%), regulatory gene classes contain no retained paralogy relations within them, reflecting either novel regulation or regulatory specialization between paralogues (Supplementary Figs 7f, h and 9d, e). Transcriptional modules exhibit an intermediate behaviour, with 26% of paralogous gene pairs having migrated between modules, both within and between the major branches of the module hierarchy (Supplementary Figs 7d, 8 and 9f).

Our analysis shows that paralogues diversify most frequently at the level of regulation, less frequently through changes in their cellular component, biological process or molecular interactions, and rarely in biochemical function. Although some of these differences may stem from variations in the quality and resolution of available annotations, multiple functional and regulatory data sources support this broad distinction. This highlights inherent limitations of gene duplication in accomplishing molecular innovation. It also emphasizes the overwhelming importance of regulatory divergence in driving functional divergence and reconfiguring molecular systems after duplication^{5,8}.

Coordinated migration of multiple paralogues is rare

When several genes in a class are duplicated they can either migrate in a coordinated manner resulting in two paralogous classes (Supplementary Fig. 12) or disperse into different classes (Fig. 5b). We expect coordinated migration after simultaneous duplications (for example, WGD⁴). To test this, we counted the number of paralogous gene pairs that connect each pair of gene classes (Supplementary Fig. 9).

Surprisingly, coordinated migration is rare: gene classes (functional, regulatory, or transcriptional) rarely share more than one or two paralogous relations, regardless of the overall proportion of paralogues retained (Supplementary Fig. 9). The few observed paralogous classes are very small and formed gradually (see, for example, Supplementary Fig. 12). Thus, paralogues overwhelmingly disperse as individuals rather than migrate in a coordinated fashion. Although theory suggests⁴ that paralogous classes might form in a single concerted step (such as after a WGD), we observed little evidence of this here.

The same patterns of retention, migration and interaction are observed even among paralogues derived only from the WGD (Supplementary Fig. 8b, data not shown) and validated by manual curation³³. Thus, while the WGD allows qualitatively different gene duplications, the subsequent patterns of innovation (or lack thereof) follow the same general trajectory for both WGD and non-WGD paralogues.

Conclusions

We set out to uncover the evolutionary potential and constraints of gene duplication and loss. We created a rich resource of evolutionary history in fungi and compared these evolutionary patterns with a wealth of functional and genomics data for *S. cerevisiae*, to uncover the principles that govern copy number changes in Ascomycota.

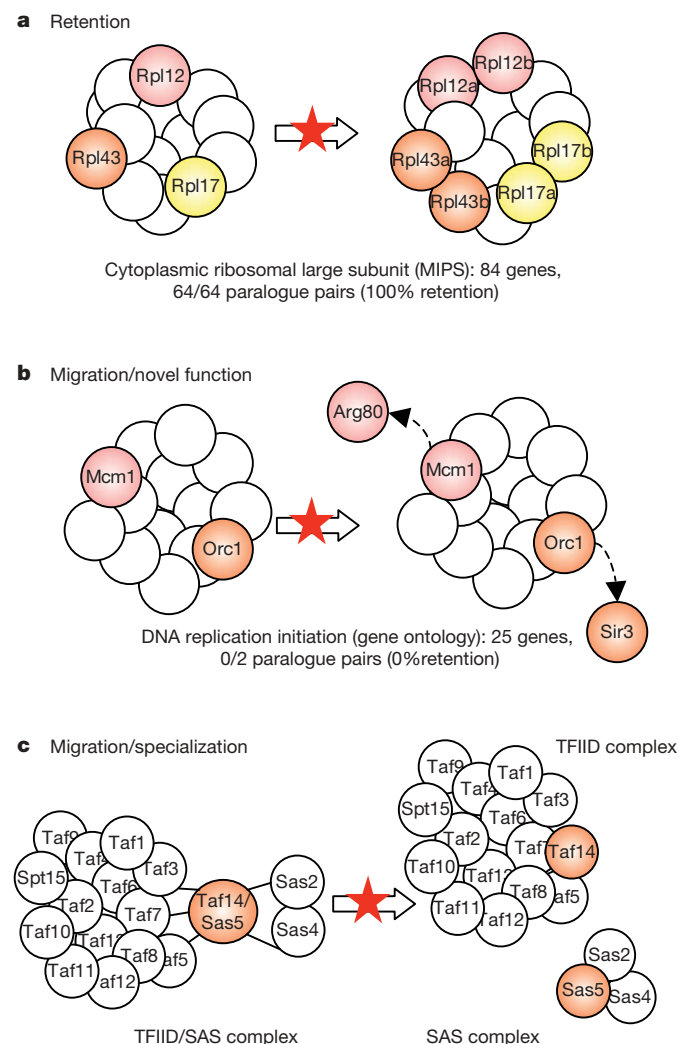


Figure 5 | Functional conservation and innovation of paralogues in classes and networks. Shown is the reconstructed functional history of several paralogous pairs of *S. cerevisiae* genes (circles with gene name). After duplication (arrow), paralogous gene pairs can be either retained within the same class (a), migrate to assume new functions (b), or specialize into distinct functions, resulting in modularization (c).

What is the contribution of gene duplication to system evolution and how does it affect the modularity of molecular systems? Earlier studies suggested that paralogous modules form in massive duplication events⁴, but we found that paralogous modules are rare, even after a WGD. An alternative mechanism is suggested by the fact that many paralogous pairs genetically interact with each other despite having no shared physical interactions (Supplementary Fig. 11c). This may indicate a partial 'division of labour' (subfunctionalization) between two paralogous proteins that become physically or temporally separated⁶. Such specialization could modularize a molecular network by severing links within a network when duplicating a node⁵⁰. For example, a single ancestral gene may have participated in two distinct complexes (Fig. 5c). If each of the paralogues specialized to optimally perform one of the functions and interact only with the members of one of the complexes, the resulting network will be more modular. Thus, increasing gene copy number may eventually simplify a system rather than making it more complex. Modularization could relax opposing constraints on a single component and thus set in motion further specialization and refinement¹¹. As our knowledge of molecular and genetic networks improves, further studies can systematically assess this possibility.

Principles similar to those described here may apply to copy number variation in metazoan genomes. For example, evidence from vertebrates and *Arabidopsis* suggests that genes encoding signalling molecules and transcription factors are duplicated in WGD events, but rarely otherwise^{3,7,9}. The reconstruction algorithm and analytical framework here make it possible to test such hypotheses in other taxa, and will facilitate other novel studies of the evolution of genes, genomes and systems.

METHODS SUMMARY

Orthology assignment and gene tree reconstruction. SYNERGY assigns orthologies in a step-wise, bottom-up fashion, solving it for each ancestral node in a species tree, starting at the leaves and concluding at the root. At each stage, SYNERGY first clusters together the genes or orthogroups from previous stages that share significant sequence similarity into new putative orthogroups (Fig. 1a). It then constructs a phylogeny of these intermediate orthogroups (Fig. 1b) using a modified neighbour-joining procedure based on the amino acid similarities that have been pre-computed and shared synteny scores. The sub-trees from each stage are based on the reconstructions at earlier stages. Each tree is rooted using a score based on sequence similarity, conserved synteny, and the inferred number of duplications and losses. Trees that invoke fewer unlikely duplication and loss events will be favoured over those that incur many such events. If the rooted tree indicates that all the orthogroups (genes) in that tree descended from a single hypothetical gene at the current stage, the cluster is defined as an orthogroup along with its tree (Fig. 1b, left). Otherwise, the orthogroup is partitioned by removing the inferred root of the gene tree (Fig. 1b, right). This process may be repeated until each orthogroup consists of genes that share a single common ancestor at the current level of reconstruction. Thus, after each stage, a complete orthology assignment and gene tree reconstruction for the genes below that node has been made. These are used as the input to subsequent stages at higher nodes in the species tree. When this procedure is completed at the root of the species tree, the genes for all species have been assigned to their orthogroups and placed in their respective trees.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 March; accepted 20 July 2007.

- Ohno, S. *Evolution by Gene Duplication* (Allen and Unwin, London, 1970).
- Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
- Blomme, T. et al. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol.* **7**, R43 (2006).
- Freeling, M. & Thomas, B. C. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.* **16**, 805–814 (2006).
- Gu, Z., Rifkin, S. A., White, K. P. & Li, W. H. Duplicate genes increase gene expression diversity within and between species. *Nature Genet.* **36**, 577–579 (2004).
- Kafri, R., Bar-Even, A. & Pilpel, Y. Transcription control reprogramming in genetic backup circuits. *Nature Genet.* **37**, 295–299 (2005).

- Maere, S. et al. Modeling gene and genome duplications in eukaryotes. *Proc. Natl Acad. Sci. USA* **102**, 5454–5459 (2005).
- Makova, K. D. & Li, W. H. Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Res.* **13**, 1638–1645 (2003).
- Papp, B., Pal, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**, 194–197 (2003).
- Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S. & Wolfe, K. H. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440**, 341–345 (2006).
- He, X. & Zhang, J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**, 1157–1164 (2005).
- Hong, E. L. B. R. et al. *Saccharomyces Genome Database* (<http://www.yeastgenome.org>) (2005).
- Arnaud, M. B. C. M. et al. *Candida Genome Database* (<http://www.candidagenome.org>) (2006).
- Dietrich, F. S. et al. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304**, 304–307 (2004).
- Dujon, B. et al. Genome evolution in yeasts. *Nature* **430**, 35–44 (2004).
- Galagan, J. E. et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**, 1105–1115 (2005).
- Fusarium graminearum* Sequencing Project. (<http://www.broad.mit.edu>) (Broad Institute of Harvard and MIT, 2003).
- Dean, R. A. et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**, 980–986 (2005).
- Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624 (2004).
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
- Wood, V. et al. The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**, 871–880 (2002).
- Arvestad, L., Berglund, A. C., Lagergren, J. & Sennblad, B. Bayesian gene/species tree reconciliation and orthology analysis using MCMC. *Bioinformatics* **19** (Suppl. 1), i7–i15 (2003).
- Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* **7**, 429–447 (2000).
- Dufayard, J. F. et al. Tree pattern matching in phylogenetic trees: automatic search for orthologs or paralogs in homologous gene sequence databases. *Bioinformatics* **21**, 2596–2603 (2005).
- Durand, D., Halldorsson, B. V. & Vernot, B. A hybrid micro-macroevolutionary approach to gene tree reconstruction. *J. Comput. Biol.* **13**, 320–335 (2006).
- Fitch, W. M. Distinguishing homologous from analogous proteins. *Syst. Zool.* **19**, 99–113 (1970).
- Jothi, R., Zotenko, E., Tasneem, A. & Przytycka, T. M. COCO-CL: hierarchical clustering of homology relations based on evolutionary correlations. *Bioinformatics* **22**, 779–788 (2006).
- Kellis, M., Patterson, N., Birren, B., Berger, B. & Lander, E. S. Methods in comparative genomics: genome correspondence, gene identification and regulatory motif discovery. *J. Comput. Biol.* **11**, 319–355 (2004).
- Li, H. et al. TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* **34**, D572–D580 (2006).
- Remm, M., Storm, C. E. & Sonnhammer, E. L. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052 (2001).
- Tatusov, R. L. et al. The COG database: an updated version includes eukaryotes. *BMC Bioinform.* **4**, article no. 41 (2003).
- Wapinski, I., Pfeffer, A., Friedman, N. & Regev, A. Automatic genome-wide reconstruction of phylogenetic gene trees. *Bioinformatics*. doi:10.1093/bioinformatics/bmt193 (2007).
- Byrne, K. P. & Wolfe, K. H. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* **15**, 1456–1461 (2005).
- Dudley, A. M., Janse, D. M., Tanay, A., Shamir, R. & Church, G. M. A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol. Syst. Biol.* **1**, 2005.0001 (2005).
- Tzung, K. W. et al. Genomic evidence for a complete sexual cycle in *Candida albicans*. *Proc. Natl Acad. Sci. USA* **98**, 3249–3253 (2001).
- Ashburner, M. et al. Gene ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
- Segal, E., Friedman, N., Kaminski, N., Regev, A. & Koller, D. From signatures to models: understanding cancer using microarrays. *Nature Genet.* **37** (Suppl.), S38–S45 (2005).
- Gavin, A. C. et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
- Deutschbauer, A. M. et al. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* **169**, 1915–1925 (2005).
- Hughes, T. R. et al. Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).

41. Newman, J. R. *et al.* Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
42. Huisinga, K. L. & Pugh, B. F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol. Cell* **13**, 573–585 (2004).
43. Tirosh, I., Weinberger, A., Carmi, M. & Barkai, N. A genetic signature of interspecies variations in gene expression. *Nature Genet.* **38**, 830–834 (2006).
44. Sopko, R. *et al.* Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* **21**, 319–330 (2006).
45. Davis, J. C. & Petrov, D. A. Do disparate mechanisms of duplication add similar genes to the genome? *Trends Genet.* **21**, 548–551 (2005).
46. Regul, T. *et al.* Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. *J. Biol.* **5**, article no. 11 (2006).
47. Kafri, R., Levy, M. & Pilpel, Y. The regulatory utilization of genetic redundancy through responsive backup circuits. *Proc. Natl Acad. Sci. USA* **103**, 11653–11658 (2006).
48. Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
49. Gerber, A. P., Herschlag, D. & Brown, P. O. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol.* **2**, article no. E79 (2004).
50. Force, A. *et al.* The origin of subfunctions and modular gene regulation. *Genetics* **170**, 433–446 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A.R. was supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and by NIGMS. N.F. was supported by the Israel Science Foundation. We thank E. S. Lander for discussions and D. Peer, A. Tanay and O. Rando for their comments on previous drafts of this manuscript. We are also grateful to the members of the FAS Center and the Broad Institute for their scientific and technical support, especially A. Daneau, M. Ethier and B. Mantenuto.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.R. (aregev@broad.mit.edu).

METHODS

Pre-processing for orthogroup reconstruction. An exhaustive sequence similarity search between all protein sequences was performed using the FASTA sequence alignment tool⁵¹. FASTA identifies pairs of sequence segments with significant similarity and subsequently performs a full alignment between them, thus returning a single hit per pair of similar genes. All hits whose expectation value *E* was below 0.1 were subsequently treated as potential homologues. This lenient threshold allows many homology relations to be considered. We then computed the amino acid similarity between each pair of potential homologues using the substitution model of ref. 52.

The SYNERGY procedure. SYNERGY assigns orthologies by tracing all genes in the species below a given node in the species tree to their reconstructed ancestral genes on the basis of sequence similarity and shared gene order³². Briefly, beginning at the leaves of the tree and traversing backwards chronologically, sets of orthogroups for two daughter species that share significant sequence similarity are assembled into candidate orthogroups at the stage pertaining to each internal node. (The orthogroups for the daughter species were identified at the previous stages; for an extant species each gene is a singleton orthogroup.) The candidate orthologues are then positioned on a phylogenetic gene tree using a modified neighbour-joining procedure that reconstructs the phylogeny of orthogroups at that stage using the amino acid similarities that have been pre-computed and shared synteny scores⁵³. Synteny scores are computed by counting the fraction of neighbouring genes that are assigned to the same orthogroup according to the provisional orthology assignments. The sub-trees joined at each stage are based on the reconstruction at earlier stages. Each tree is rooted using a combined score based on sequence similarity, conserved synteny, and the inferred number of duplications and losses. Trees that invoke fewer unlikely duplication and loss events will be favoured over those that incur many such events. If the rooted tree determines that a candidate orthogroup's genes are not all descended from a single ancestral gene, the orthogroup is partitioned to two candidate orthogroups by removing the inferred root of the gene tree. This process may be repeated on the two new trees until each tree represents genes that share a single common ancestor at the current level of reconstruction. The orthology assignments and trees obtained at each stage of the algorithm are used as the input to the next stages, involving higher nodes in the species tree. The interim trees are also used to compute the distances between orthogroups by using the neighbour-joining distance update rule when nodes on the tree are joined. When this procedure is completed at the root of the species tree, the genes for all species have been assigned to their orthogroups and placed in their respective trees. Full details of the procedure are available in the companion technical manuscript³².

Bootstrap evaluation. To assess the sensitivity of SYNERGY's orthology assignments to both the choice species and their protein-coding gene predictions, two bootstrap-based confidence measures were calculated. A 'branch-bootstrap' was performed by systematically removing branches from the species tree and counting the number of orthology assignments that changed. A 'gene-bootstrap' was similarly executed by hiding each predicted open reading frame in the data set with a fixed probability of 0.2. For each bootstrap run, a predicted orthogroup was validated for the number of orthology assignments that were not designated during that run ('soundness'), and the proportion of the original orthologous pairs among the new predictions ('completeness'). For each pair of orthologues within an orthogroup from a bootstrap run, an assignment was considered to be inconsistent either if the pair was assigned as orthologues at a point in the species tree that deviated from the original assignment or if they were not originally assigned as orthologues at all. The average number of consistent assignments across each experiment was used as the branch- and gene-bootstrap confidence measures for an orthogroup's soundness and completeness. A detailed description of this procedure is given in Supplementary Note 1.

Genome sequences. The predicted protein sequences and their genomic locations were downloaded from the following sources. From the *Saccharomyces* Genome Database¹²: *S. cerevisiae* (downloaded July 2005), *S. paradoxus* (accession number AABY01000000), *S. mikatae* (accession number AABZ01000000), *S. bayanus* (accession number AACA01000000) and *S. castellii* (accession number AACF00000000). From Génolevures¹⁵: *C. glabrata* (accession number CR380947–CR380959), *K. lactis* (accession number CR382121–CR382126), *D. hansenii* (accession number CR382133–CR382139) and *Y. lipolytica* (accession number CR382127–CR382132). From the *Ashbya* Genome Database¹⁴: *A. gossypii* (accession number AE016814–AE01682). From Kellis et al.¹⁹: *K. waltii* (accession number AADM01000000). From the *Candida* Genome Database¹³: *C. albicans* (accession number AACQ00000000). From The Broad Institute Fungal Genome Initiative (<http://www.broad.mit.edu/annotation/fgi/>): *Aspergillus nidulans* (accession number AACD00000000), *Fusarium graminearum* (accession number AACM00000000), *Magnaporthe grisea* (accession number AACU00000000), *Neurospora crassa* (accession number

AABX00000000) and *Cryptococcus neoformans* (accession number AACO01000000). From PombeDB²¹: *S. pombe* (accession numbers CU329670–CU329672).

Species tree. The species tree topology representing the phylogenetic relations between the Ascomycota fungi was compiled from several sources^{33,54,55} and subsequently validated using the results from our orthogroup assembly by concatenating the multiple sequence alignments from 50 sampled uniform orthogroups and obtaining the maximum-likelihood tree topology using the Phylip package's default parameters⁵⁶. This sampling procedure was repeated ten times, with the same tree topology resulting each time. A manual correction to this topology was introduced as previously described³³, swapping the locations of *S. castellii* and *C. glabrata* owing to chromosome-based evidence suggesting that *C. glabrata* is in fact more closely related to the *Saccharomyces* sensu stricto clade. Branch lengths for the tree presented in Fig. 1a were estimated using a maximum-likelihood approach using multiple site rate variation with the default parameters⁵⁷. Branch lengths were not used in any of the subsequent analyses.

Functional gene classes. We compiled a total of 3,395 gene classes, obtained as follows: 1,794 from the Gene Ontology³⁶ hierarchy, 87 from the Kyoto Encyclopedia of Genes and Genomes⁵⁸ (KEGG), 107 from the BioCyc database⁵⁹, 1,022 from the MIPS database of manually curated protein complexes⁶⁰, 310 from a data set describing the targets genes bound by various transcription factors⁴⁸, 70 from a data set describing the target genes harbouring a given *cis*-regulatory element in their promoter⁴⁸, and 5 from a data set describing the targets of the RNA binding proteins from the PUF family⁴⁹. These classes were used for constructing the transcription module hierarchy (see below). In addition, the following gene classes were included in various analyses, but were not included in the construction of the module hierarchy: genes controlled by the SAGA and/or TFIID transcription complexes⁴², genes with and without TATA box control⁴³, genes with large levels of expression variation between yeast species⁴³, genes with high and low levels of noise in protein abundance⁴¹, haploinsufficient genes³⁹, genes whose overexpression reduces fitness⁴⁴, and genes belonging to complex cores, attachments and modules based on high throughput assays³⁸.

Functional orthogroup classes. We used the aforementioned *S. cerevisiae*-based gene classes to define orthogroup classes by projecting the *S. cerevisiae* annotations onto the orthogroups containing *S. cerevisiae* genes.

***S. cerevisiae* protein interaction networks.** We constructed separate biochemical and genetic protein interaction networks using both manually curated and high-throughput data sources⁴⁶.

Defining orthogroup copy number variation profiles. To measure the changes of gene copy number through either duplication or loss events, we assigned a copy number variation profile to each orthogroup. The profile is defined by inspecting the extended phylogenetic profiles belonging to an orthogroup, and subtracting the number of losses observed at each index in the species tree from the number of duplications. We increment this copy number variation profile at the last common ancestor identified for the orthogroup, indicating its 'age'.

Coherence of copy number variation profiles. To assess the coherence in gene copy number variation across a class of orthogroups, we first calculated the class' centroid extended copy number variation profile (ECVP) by averaging the ECVPs from all the orthogroups belonging to a class. This centroid is then applied to estimate the degree of deviation between the orthogroups belonging to a class by summing the L1 distance from each of the class' orthogroups to it. We compare this deviation to that of 10,000 randomly assigned orthogroup classes, each containing the same number of ECVPs. The fraction of times the deviation is equal to or less than that of the orthogroup class is the measure (*P*-value) we use to evaluate the coherence of that class. Copy number variation occurs at each individual branch of the species tree, so we similarly define a coherence profile for an orthogroup class by evaluating the copy number variation coherence for each position along the species tree.

Copy number variation in protein interaction networks. To test the relation between proximity in protein interaction networks and similarity in copy number variation, we first computed the difference (using the L1 distance) between the ECVPs for each pair of proteins in the network, ignoring pairs that belong to the same orthogroup (hence sharing the same profile). Next we averaged these differences among all proteins within a given radius in the network. To determine whether these averages were significant, we repeated this computation by shuffling profile assignments to proteins in the network 1,000 times, obtaining the expected range of average differences between pairs of proteins in the network for each radius.

Statistical benchmark of orthogroup evolutionary history. To benchmark the volatility in gene copy number for each orthogroup, we used the estimated rates of duplications and losses along each branch of the species tree to calculate the log-probability of the observed number of such events in each orthogroup, assuming that they occur according to a standard Poisson distribution. This

statistic is used as a measure of volatility for each orthogroup. We compare this volatility metric to those of 10,000 hypothetical orthogroups with randomly generated duplications and losses (based on the empirical rates). In our analysis we label those orthogroups the volatility of which deviates more than three standard deviations from the mean of the random distribution as being significantly volatile.

Statistical enrichment tests. To identify the annotations in which each orthogroup class was significantly enriched, we projected the class' annotations as described above, calculated the fraction of orthogroups from that class that contained a given annotation and used the hypergeometric distribution to calculate a *P*-value for this fraction (compared with the null hypothesis of choosing the same number orthogroups at random). We corrected for multiple tests using the false discovery rate correction with a 0.01% false rate.

Construction of *S. cerevisiae* transcription module map. We constructed the hierarchy of transcriptional modules following the procedure presented by Segal *et al.*³⁷. We applied this procedure to a yeast data set and followed it by manually selecting which modules to use in the hierarchy. We note that although it is not highlighted in their manuscript, this method creates modules in a hierarchical fashion. Full technical details on the construction of the map are at the end of this section.

Estimating functional divergence of paralogous genes in hierarchical annotation data. We estimated the functional divergence between a pair of paralogous genes by considering the most specific gene class in the annotation hierarchy that each gene was assigned to (for gene ontology we ignored all assignments that were attributed only to computational sequence analysis). We regarded a pair of genes as functionally diverged only if both genes are assigned to at least one annotation class and they are not both assigned to the most specific annotations of either of the two genes.

Computing degree of conserved interactions between paralogues. We used two statistics to compute the degree of conserved interactions between pairs of paralogous proteins. The first was simply the fraction of shared interactions between both proteins. For this we counted the number of interactions each protein takes part in (a_1 and a_2 for proteins 1 and 2, respectively) and the number of interactions they both share (s). The fraction of shared interactions is thus:

$$f = s / [\min(a_1, a_2)]$$

We also used the subfunctionalization index (I_{sf}) as previously described¹¹ to characterize how diverged a pair of paralogues' interactions are. This is calculated as:

$$I_{sf} = 1 - (s + |a_1 - a_2|) / t$$

where s is as above and t is equal to $(a_1 + a_2 - s)$. This statistic gives a reasonable estimate of the degree of subfunctionalization in the absence of neofunctionalization, because subfunctionalization would reduce the number of shared interactions. This measure considers the proportion of ancestral interactions that are no longer shared between the paralogues and the extent of subfunctionalization for these interactions.

Estimating significance of shared protein interaction neighbourhood. To estimate the significance of the shared protein interaction neighbourhood between pairs of paralogues, we first calculated the degree of conserved interactions and compared this to the degree of shared interactions between the two paralogues in a degree-preserving randomized network, obtained by swapping edges between random pairs of nodes 10^6 times. We repeated this procedure 10,000 times, and assigned a *P*-value to the shared protein interaction neighbourhood of a pair of proteins according to the number of times the fraction of shared interacting partners between paralogues is equal to or greater than the fraction in the real network.

Homology searches. To identify putative homologies between orthogroups and the *C. neoformans* genome, we first constructed a tree-assisted multiple sequence alignment from each orthogroup's protein sequences using their reconstructed gene trees and the MUSCLE alignment software⁶¹. We then constructed a sequence profile from this alignment, and executed a homologous protein search using the HMMSEARCH profile search software, employing an *E*-value cutoff of 1.0 (ref. 62).

DNA microarray data set. We compiled a collection of 1,216 previously published microarray experiments (Supplementary Table 6). We normalized the expression of each gene g in every data set separately as in ref. 37. For data sets generated using Affymetrix chips, we first take the log (base 2) of g 's expression value in each array (truncating expression values that are below ten). For data sets generated using spotted complementary DNA chips, we use the log-ratio (base 2) between the measured sample and the control sample. In both types of data sets, we subsequently normalize the (log) expression value of gene g in each

array to its average expression in all the arrays in the same data set by subtracting its average in that data set from each of its expression measurements. After this normalization, the mean value of a gene within each data set is zero.

Details of transcriptional module hierarchy construction. The transcription module hierarchy was constructed in a step-wise process as in ref. 37.

Step 1: Identifying arrays where gene classes significantly change in expression. To identify the arrays where each gene class is significantly induced (or repressed), we defined the induced (or repressed) genes in each array to be those genes whose change in expression is greater (less) than twofold. For each gene class and each array, we calculated the fraction of genes from that class that are induced (or repressed) in that array, and used the hypergeometric distribution to calculate a *P*-value for this fraction (compared to the null hypothesis of choosing the same number genes at random). We corrected for multiple tests using the false discovery rate correction with a 1% false rate.

Step 2: Identification of gene class clusters. We performed (bottom-up) hierarchical clustering of the gene classes in the matrix of all significant array-gene class pairs. We manually selected a hierarchy of gene class clusters corresponding to the cluster boundaries defined automatically, and assigned a biologically meaningful name to each cluster. We obtained a total of 163 such gene class clusters (excluding the root node). The transcriptional modules were defined from the genes in those gene class clusters (below), and organized according to the same hierarchy.

Step 3: Testing consistency of a gene with expression of a gene class. Given a class of genes G and a gene g , we test whether the expression of g is consistent with the significant changes in the expression of G using the following procedure. We first identify the subsets of arrays I and R where G is significantly induced and repressed, respectively. We then measure the extent in which the expression of g changes by more (or less) than twofold in arrays in I (or R) with the score

$$\text{Score}(g) = \sum_{\{a \in I \mid g \text{ is induced in } a\}} -\log(p_a) + \sum_{\{a \in R \mid g \text{ is repressed in } a\}} -\log(p_a)$$

where p_a is the fraction of genes in the array a that are induced (or repressed) by more than twofold for arrays in I (or in R). This score assigns more weight to induction in arrays where there are fewer induced genes (and respectively for repression).

We evaluate the significance of $\text{Score}(g)$ with respect to the null hypothesis in which the genes in each array are randomly permuted. Under this null hypothesis, $\text{Score}(g)$ is the sum of independent binary random variables, one for each array in I and R . The random variable corresponding to array a attains the value $-\log(p_a)$ with probability p_a and the value of 0 with probability $1 - p_a$. Because $\text{Score}(g)$ in this model is a sum of independent random variables, its mean μ and variance σ^2 are the sum of the means and variances, respectively, of these variables, and can be computed analytically:

$$\mu = \sum_{a \in I \cup R} -p_a \log p_a$$

$$\sigma^2 = \sum_{a \in I \cup R} p_a (1 - p_a) \log^2 p_a$$

By the central limit theorem, the distribution of $\text{Score}(g)$ under the null hypothesis can be closely approximated by a gaussian distribution with mean μ and variance σ^2 . We use standard methods for computing the tail probability of a gaussian distribution to compute the probability of attaining a score as large as the observed score under the null hypothesis.

Step 4: Deriving modules from clusters of gene classes. For each cluster of gene classes, we define G to be the union of the gene classes in the cluster. We then test each gene in G for consistency (as described above). The resulting module consists of genes the expression of which is significantly consistent with the expression of the gene class (after false discovery rate correction for multiple hypotheses using a 0.01% false rate).

51. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448 (1988).
52. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282 (1992).
53. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
54. Kurtzman, C. P. & Robnett, C. J. Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. *FEMS Yeast Res.* **3**, 417–432 (2003).
55. Kuramae, E. E., Robert, V., Snel, B. & Boekhout, T. Conflicting phylogenetic position of *Schizosaccharomyces pombe*. *Genomics* **88**, 387–393 (2006).
56. Felsenstein, J. PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166 (1989).

57. Ninio, M., Privman, E., Pupko, T. & Friedman, N. Phylogeny reconstruction: increasing the accuracy of pairwise distance estimation using Bayesian inference of evolutionary rates. *Bioinformatics* **23**, e136–e141 (2007).
58. Kanehisa, M. A database for post-genome analysis. *Trends Genet.* **13**, 375–376 (1997).
59. Karp, P. D. *et al.* Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res.* **33**, 6083–6089 (2005).
60. Mewes, H. W. *et al.* MIPS: analysis and annotation of proteins from whole genomes in 2005. *Nucleic Acids Res.* **34**, D169–D172 (2006).
61. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
62. Eddy, S. HMMER: Hidden Markov models for sequence profile analysis. (<http://hmmer.janelia.org/>) (2003).

ARTICLES

Glycosphingolipid synthesis requires FAPP2 transfer of glucosylceramide

Giovanni D'Angelo¹, Elena Polishchuk¹, Giuseppe Di Tullio¹, Michele Santoro¹, Antonella Di Campi¹, Anna Godi^{1†}, Gun West², Jacek Bielawski³, Chia-Chen Chuang⁴, Aarnoud C. van der Spoel⁴, Frances M. Platt⁴, Yusuf A. Hannun³, Roman Polishchuk¹, Peter Mattjus² & Maria Antonietta De Matteis¹

The molecular machinery responsible for the generation of transport carriers moving from the Golgi complex to the plasma membrane relies on a tight interplay between proteins and lipids. Among the lipid-binding proteins of this machinery, we previously identified the four-phosphate adaptor protein FAPP2, the pleckstrin homology domain of which binds phosphatidylinositol 4-phosphate and the small GTPase ARF1. FAPP2 also possesses a glycolipid-transfer-protein homology domain. Here we show that human FAPP2 is a glucosylceramide-transfer protein that has a pivotal role in the synthesis of complex glycosphingolipids, key structural and signalling components of the plasma membrane. The requirement for FAPP2 makes the whole glycosphingolipid synthetic pathway sensitive to regulation by phosphatidylinositol 4-phosphate and ARF1. Thus, by coupling the synthesis of glycosphingolipids with their export to the cell surface, FAPP2 emerges as crucial in determining the lipid identity and composition of the plasma membrane.

The Golgi complex (GC) is responsible for processing, trafficking and sorting of proteins and lipids synthesized in the endoplasmic reticulum (ER). The *trans*-Golgi network (TGN) is the major sorting compartment of the GC, where cargoes are packaged into distinct carriers for transport to their final destinations. Our present understanding of the formation of carriers moving from the TGN to the plasma membrane (PM) is limited^{1,2}. Several proteins involved in their biogenesis act through a tight interplay with lipids^{1–3}; among these we identified FAPP1 and FAPP2 (ref. 4). The FAPPs are recruited to the TGN through their pleckstrin homology (PH) domain, which binds phosphatidylinositol 4-phosphate (PtdIns4P), and the small GTPase ARF1. The FAPPs are involved specifically in transport from the TGN to the PM^{4,5} and control cilium formation in polarized cells⁶.

The mechanisms of action of the FAPPs remain to be defined. FAPP2 has sequence similarity to glycolipid-transfer protein (GLTP)^{4,7}, a protein able to mediate the intermembrane transfer of glycosphingolipids (GSLs) *in vitro*⁸. GSLs are important components of the PM, with key roles in cell signalling, adhesion, proliferation and differentiation⁹.

GSL synthesis starts in the ER from ceramide. Ceramide is glycosylated to glucosylceramide (GlcCer) on the cytosolic leaflet of *cis*-Golgi membranes by GlcCer synthase (GCS)¹⁰. GlcCer has then to be transported to late-Golgi compartments and translocated to the luminal leaflet, because all further glycosylation steps leading to complex GSLs take place in the lumen of these compartments^{11,12}. Thus, its presence on the cytosolic face of the GC gives GlcCer the correct topology to be 'intercepted' by cytosolic proteins that can bind to and/or transfer it. The transfer of GlcCer by cytosolic GLTPs would permit the non-vesicular (that is, non-membrane-bound) transport of GlcCer, as opposed to or in addition to its vesicular transport (by means of membrane flow).

Here we show that FAPP2 is required for the synthesis of complex GSLs because it mediates the non-vesicular transport of GlcCer to

distal Golgi compartments. This puts the GSL biosynthetic pathway under the control of PtdIns4P and the small GTPase ARF1, the two known regulators of FAPP2.

FAPP2 is a GlcCer-transfer protein

Despite the limited sequence similarity of the GLTP-homology domain of FAPP2 to GLTP itself, the residues required for the glycolipid transfer activity of GLTP¹³ are all conserved in FAPP2 (Supplementary Fig. 1). Furthermore, homology modelling of FAPP2 on the structure of GLTP¹³ produces reliable and significant overlap (Fig. 1a). These considerations prompted us to investigate whether FAPP2 is indeed a glycolipid-transfer protein.

Figure 1b shows that FAPP2 can transfer dodecanoyl-dipyrromethene boron difluoride-GlcCer (C12-BODIPY-GlcCer) between 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) vesicles *in vitro*¹⁰, whereas a mutated FAPP2 designed on an inactive mutant of GLTP¹³ (FAPP2-W407A) cannot (Fig. 1a–c).

To define the lipid specificity of FAPP2, its ability to transfer sphingomyelin (SM), phosphatidylcholine and ceramide was examined. As shown in Fig. 1b, c, FAPP2 did not transfer SM, phosphatidylcholine or ceramide. Because the FAPP PH domain binds PtdIns4P and ARF1 (ref. 4), we assessed whether these two PH-domain ligands affect the GlcCer transfer activity of FAPP2. The addition of PtdIns4P and ARF1 to acceptor (but not to donor) vesicles increased the rate of GlcCer transfer by FAPP2 (Fig. 1d).

The FAPP2 lipid-transfer activity for longer and unmodified acyl-chain glycolipids was assessed with the use of [¹⁴C]stearoyl-GlcCer. As shown in Fig. 1e, the addition of FAPP2 and GLTP produced substantial transfer of [¹⁴C]stearoyl-GlcCer to acceptor vesicles.

These data demonstrate that FAPP2 acts *in vitro* as a GLTP, transferring both BODIPY-labelled and unmodified long-acyl-chain GlcCer.

¹Department of Cell Biology and Oncology, Consorzio Mario Negri Sud, Via Nazionale 8/A, 66030 Santa Maria Imbaro, Chieti, Italy. ²Department of Biochemistry and Pharmacy, Åbo Akademi University, Artillerigatan 6 A III, BioCity FI-20520 Turku, Finland. ³Department of Biochemistry and Molecular Biology, Medical University of South Carolina, 173 Ashley Avenue, Charleston, South Carolina 29425, USA. ⁴Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, UK. [†]Present address: Section of Cell and Molecular Biology, Institute of Cancer Research Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK.

We then examined the ability of FAPP2 to transfer C12-BODIPY-GlcCer in permeabilized cells (see Methods). Here C12-BODIPY-GlcCer was delivered by FAPP2 preferentially to the perinuclear Golgi area (Fig. 1f). The FAPP2-W407A mutant, inactive in the GlcCer transfer assay *in vitro* (Fig. 1b, c), did not deliver GlcCer to the GC (Fig. 1f). This FAPP2-mediated delivery of GlcCer to the GC is dependent on ARF1 and PtdIns4P, because it was abolished when ARF or phosphatidylinositol-4-OH kinases (PI(4)Ks) were inactivated by brefeldin A (BFA)¹⁴ or phenylarsineoxide (PAO)¹¹, respectively (Fig. 1f). GLTP itself showed no organelle preference for the delivery of C12-BODIPY-GlcCer to cell membranes (Fig. 1f).

Last, to analyse the role of FAPP2 in influencing the distribution of GlcCer *in vivo*, mapping of GlcCer by immunoelectron microscopy was performed with a specific anti-GlcCer antibody¹² (see Supplementary Information and Supplementary Fig. 2) in control cells and in cells knocked down for FAPP2 (Supplementary Fig. 3). Whereas in control cells GlcCer was concentrated at the TGN (Fig. 1g), in cells

knocked down for FAPP2 it was mainly on the *cis* pole of the Golgi stacks (Fig. 1g).

These results indicate that at steady state the transport of GlcCer from *cis*-Golgi to *trans*-Golgi is very efficient in control cells. In contrast, in the absence of FAPP2, GlcCer persists longer at its site of synthesis in the *cis*-Golgi.

Taken together, these data show that FAPP2 is a GlcCer-transfer protein that controls the transfer of GlcCer from *cis*-Golgi to *trans*-Golgi, a key step in GSL synthesis.

FAPP2 is required for GSL synthesis

The impact of FAPP2 knockdown on GSL synthesis was analysed in HeLa cells (and in HT29 and COS7 cells; not shown) in which the sphingolipids had been labelled with [³H]sphingosine (Fig. 2 and Supplementary Fig. 4).

FAPP2 knockdown markedly impaired the synthesis of GSLs (70% average inhibition) (Fig. 2a, b): the levels of all GSLs downstream

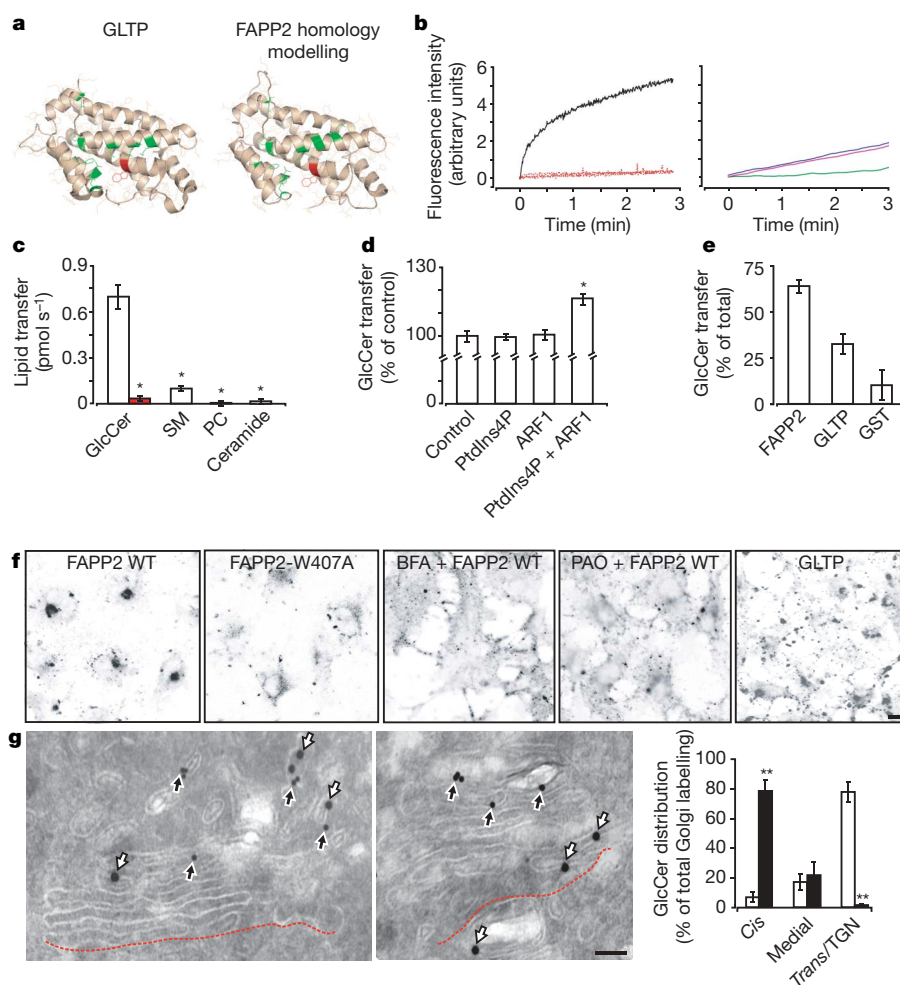


Figure 1 | FAPP2 is a GlcCer-transfer protein. **a**, Homology modelling of the GLTP homology domain of FAPP2 (see Supplementary Information). Red, tryptophan residue required for glycolipid transfer activity (GLTP, Trp 96; FAPP2, Trp 407). Green, residues important for glycolipid transfer activity in GLTP. **b**, Lipid transfer *in vitro* by fluorescence resonance energy transfer. Left: C12-BODIPY-GlcCer with wild-type FAPP2 (black) and FAPP2-W407A (red). Right: BODIPY-SM (blue), C12-BODIPY-ceramide (red) and C12-BODIPY-phosphatidylcholine (green) with wild-type FAPP2. **c**, Quantification of **b**. White columns, wild-type FAPP2; red columns, FAPP2-W407A. **d**, Effects of ARF1 and PtdIns4P on FAPP2-mediated transfer of GlcCer. Acceptor vesicles were 2 mol% PtdIns4P and/or 0.2 μ M ARF1 and 40 μ M GTP. **e**, Transfer of [¹⁴C]stearoyl-GlcCer *in vitro*. GST, glutathione S-transferase. **f**, C12-BODIPY-GlcCer transfer in permeabilized cells with 1 μ M purified wild-type (WT) FAPP2, FAPP2-W407A and GLTP

preincubated with C12-BODIPY-GlcCer. BFA (5 μ g ml⁻¹) and PAO (10 μ M) were also used as indicated. BODIPY-GlcCer fluorescence is shown. Results are representative of three independent experiments, each in duplicate. Scale bar, 10 μ m. **g**, Sub-Golgi distribution of GlcCer. Mock-treated (left) and FAPP2 knockdown (middle) HeLa cells were labelled with anti-GlcCer (15-nm gold particles, white arrows) and anti-TGN46 (10-nm gold particles, black arrows) antibodies; the red dashed line indicates a fenestrated *cis*-Golgi cisterna. Scale bar, 100 nm. Right: quantification by counting gold particles. Open columns, mock; filled columns, FAPP2 knockdown. Data in **c**, **d** and **g** are means \pm s.d. for three independent experiments, each in duplicate ($n = 3$), and in **e** are means \pm range for two independent experiments, each in triplicate ($n = 2$). Asterisks indicate statistical significance (see Methods).

of GlcCer in the biosynthetic pathway were lower in FAPP2-knockdown cells than in control cells, starting from lactosylceramide (LacCer), the first product of GlcCer metabolism, and including more complex GSLs (collectively indicated as GSLs in Fig. 2). The inhibition of GSL synthesis was maintained at all [3 H]sphingosine pulse times, from 30 min to 24 h (Fig. 2b).

With FAPP2 knockdown after a 24-h [3 H]sphingosine pulse, the steady-state levels of radiolabelled GlcCer were about double those of the mock-treated cells (Fig. 2b). However, after a 2-h [3 H]sphingosine pulse, no increase in GlcCer was seen under FAPP2 knockdown. This might be due either to the enhanced degradation of GlcCer by Glc-ceramidases^{15,16} or to its decreased synthesis.

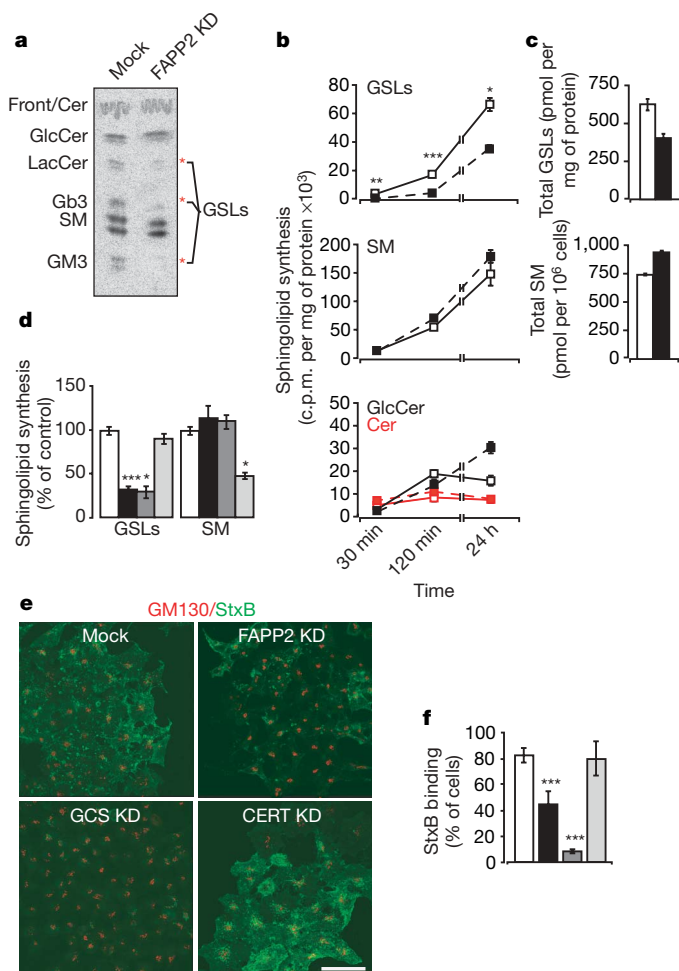


Figure 2 | FAPP2 is required for complex GSL synthesis. **a**, Representative TLC of [3 H]sphingosine-labelled lipids (120-min pulse) in mock-treated and FAPP2 knockdown (FAPP2 KD) HeLa cells. Front/Cer, irregular ceramide band migrating with fatty-acid-enriched eluant front. **b**, Quantification of radiolabelled GSLs (top), SM (middle) and GlcCer and ceramide (bottom) in mock-treated (open squares) and FAPP2 KD (filled squares) HeLa cells after [3 H]sphingosine pulses. **c**, Quantification of endogenous GSL (by HPLC) and SM levels (by LC-MS). Open columns, mock; filled columns, FAPP2 KD. **d**, Radiolabelled GSLs and SM (120-min [3 H]sphingosine pulse) in mock-treated (white columns), FAPP2 KD (black columns), GCS KD (dark grey columns) and CERT KD (light grey columns) HeLa cells. **e**, StxB immunofluorescence in mock-treated, FAPP2 KD, GCS KD and CERT KD HeLa cells. Merged images of StxB-dichlorotriazinylaminofluorescein (green) and GM130 (red) are shown. Scale bar, 50 μ m. **f**, Quantification of StxB binding in mock-treated (white columns), FAPP2 KD (black columns), GCS KD (dark grey columns) and CERT KD (light grey columns) cells. Data in **b**, **c** and **d** are means \pm s.e.m. for three independent experiments, each in duplicate ($n = 3$), and in **e** are means \pm s.d. for five independent experiments, each in duplicate ($n = 5$). Asterisks indicate statistical significance.

To distinguish between these two possibilities, we treated cells with Glc-ceramidase inhibitors (1 mM conduritol B epoxide and 5 μ M *N*-butyldeoxynojirimycin). Also under these conditions there was no significant increase in GlcCer levels in FAPP2-knockdown cells with the 2-h pulse (not shown), thus indicating a possible inhibition of GCS activity by the high local levels of GlcCer in FAPP2-knockdown cells.

SM synthesis was not inhibited by FAPP2 knockdown; indeed, SM levels were even higher than those of control cells (Fig. 2b–d). Measurement of endogenous GSLs (by high-performance liquid chromatography (HPLC)) and SM (by liquid chromatography–mass spectrometry (LC-MS)) confirmed both the decrease in GSL levels and the increase in SM levels in FAPP2-knockdown cells, compared with mock-treated cells (Fig. 2c).

As a positive control, proteins with an established role in GSL and SM synthesis, such as GCS and the ceramide-transfer protein (CERT)¹⁷, were knocked down: there was a marked decrease in GSL synthesis in GCS-knockdown cells, whereas a decrease in SM synthesis was seen with CERT knockdown (Fig. 2d and Supplementary Fig. 3).

The role of FAPP2 in controlling GSL levels was finally confirmed by assessing the binding of subunit B of Shiga toxin (StxB), which uses the GSL Gb3 as its receptor on the cell surface¹⁸. As shown in Fig. 2e, f, binding of StxB to the PM was decreased in FAPP2-knockdown cells in comparison with mock-treated cells.

Taken together, these findings show that the synthesis and steady-state levels of GSLs are under the control of FAPP2.

Uncoupling of GSL synthesis from vesicular trafficking

The finding that GSL synthesis is dependent on the activity of a cytosolic lipid-transfer protein such as FAPP2 raises the issue of the relative contributions of vesicular and non-vesicular transport to the transfer of GlcCer to the late-Golgi compartments, the site of its conversion into GSLs.

The requirement for membrane trafficking in GSL biosynthesis was previously deduced essentially from inhibitory effects of the Golgi-disrupting agent BFA on the synthesis of complex GSLs¹⁹. However, BFA exerts its effects by inactivating ARF1 and causing the release of its effectors into the cytosol, including FAPP2 (ref. 4). Thus, BFA blocks membrane-bound and non-membrane-bound transport of GlcCer; this therefore precludes its use to determine the relative contributions of these two pathways.

To block the membrane-bound transport of GlcCer selectively, two inhibitors of intra-Golgi membrane trafficking, namely dicoumarol²⁰ and the phospholipase A₂ inhibitor ONO-RS-082 (refs 21, 22), were used, which do not affect the association of ARF effectors with the GC²⁰. We compared the effects of the inhibitors on intra-Golgi transport of the reporter protein VSVG (vesicular stomatitis virus G protein) and on GSL synthesis. Whereas the transport of VSVG was severely impaired by dicoumarol, synthesis of GSL was not affected (Fig. 3a). With phospholipase A₂, inhibition by ONO-RS-082 and the specific knockdown of phospholipase A₂ α by RNA interference caused the selective breakdown of tubular connections between adjacent Golgi stacks and between adjacent cisternae of the same stack²³, and arrested the transport of newly synthesized cargoes in the *cis*-Golgi cisterna (R. Polishchuk, personal communication). Here, in ONO-RS-082-treated cells, there were no changes in GSL synthesis, despite marked inhibition of VSVG transport (Fig. 3a).

Next, we performed a knockdown of Bet3 (Supplementary Fig. 3), a component of the TRAPP complex required for membrane trafficking and maintenance of the structure of the GC²⁴. Under these conditions, GSL synthesis was not substantially affected compared with mock-treated cells, whereas transport of VSVG was inhibited (Fig. 3b).

Last, a correlative analysis of the rate and extent of GSL synthesis and of membrane flow to and through the GC was performed. HeLa

cells were treated with the protein synthesis inhibitor cycloheximide; under these conditions, fewer transport carriers were generated from the ER and arrived at the GC, the total surface area of the GC decreased, and the Golgi ribbon was reduced to individual disconnected stacks. These are all signs of quiescent Golgi stacks^{23,25}. Under these conditions, however, GSLs were synthesized at rates and to extents comparable to those under control conditions (Fig. 3b).

Taken together, these data indicate that non-vesicular transport of GlcCer by FAPP2 provides a major contribution to GSL synthesis, because this can apparently proceed at a normal rate even when membrane trafficking through the GC is severely impaired.

GSL synthesis is controlled by PtdIns4P

Because FAPP2 is targeted to the TGN through PtdIns4P (ref. 3), we investigated whether interfering with PtdIns4P at the GC has any impact on GSL synthesis. PtdIns4P is produced at the GC by PI(4)KII α and PI(4)KIII β ²⁶. We first used a general inhibitor of PI(4)Ks, PAO^{4,5,11} (Fig. 3c). PAO induced a marked inhibition of synthesis of GSLs and inhibited SM synthesis, the latter in line with the requirement for PtdIns4P for the targeting and activity of CERT^{17,26}. Next, the expression of PI(4)KII α and PI(4)KIII β was specifically knocked down with small interfering RNAs (siRNAs; Supplementary Fig. 3): this also resulted in a marked inhibition of GSL and SM synthesis (Fig. 3c), highlighting a role for phosphoinositides as controllers of sphingolipid synthesis.

GlcCer transfer by FAPP2 controls TGN-to-PM transport

Thus, FAPP2, a protein required for TGN-to-PM transport^{4,5}, is a GlcCer-transfer protein with a key function in GSL synthesis. This raises the question of whether the GlcCer transfer activity of FAPP2 is relevant for its involvement in membrane trafficking.

In agreement with our previous report⁴, TGN-to-PM transport of VSVG and release of glycosaminoglycans were impaired in FAPP2-knockdown cells, compared with mock-treated cells (Fig. 3d, e, and Supplementary Fig. 5a). In addition, FAPP2 knockdown inhibited the secretion of albumin (Supplementary Fig. 5b) and total glycoproteins (not shown) in HepG2 and HeLa cells, respectively. To evaluate the role of the GlcCer transfer activity of FAPP2 in membrane trafficking, we assessed the effects of re-expressing siRNA-resistant forms of wild-type FAPP2 and the GlcCer-transfer-dead form FAPP2-W407A (Supplementary Fig. 3) on impaired TGN-to-PM transport induced by endogenous FAPP2 knockdown. Although almost complete rescue of TGN-to-PM transport was seen in cells expressing wild-type FAPP2, only a very limited recovery was seen in cells expressing FAPP2-W407A (Fig. 3d). A single point mutant of FAPP2 that cannot interact with PtdIns4P (FAPP2-R18L; Supplementary Fig. 6) did not rescue VSVG transport in FAPP2-knockdown cells, thus confirming the importance of PtdIns4P in controlling FAPP2 activity (Fig. 3e).

This coupling of metabolic and trafficking functions of FAPP2 indicated that GSL synthesis controls TGN-to-PM transport. To specifically explore the role of GSLs in membrane trafficking at the TGN,

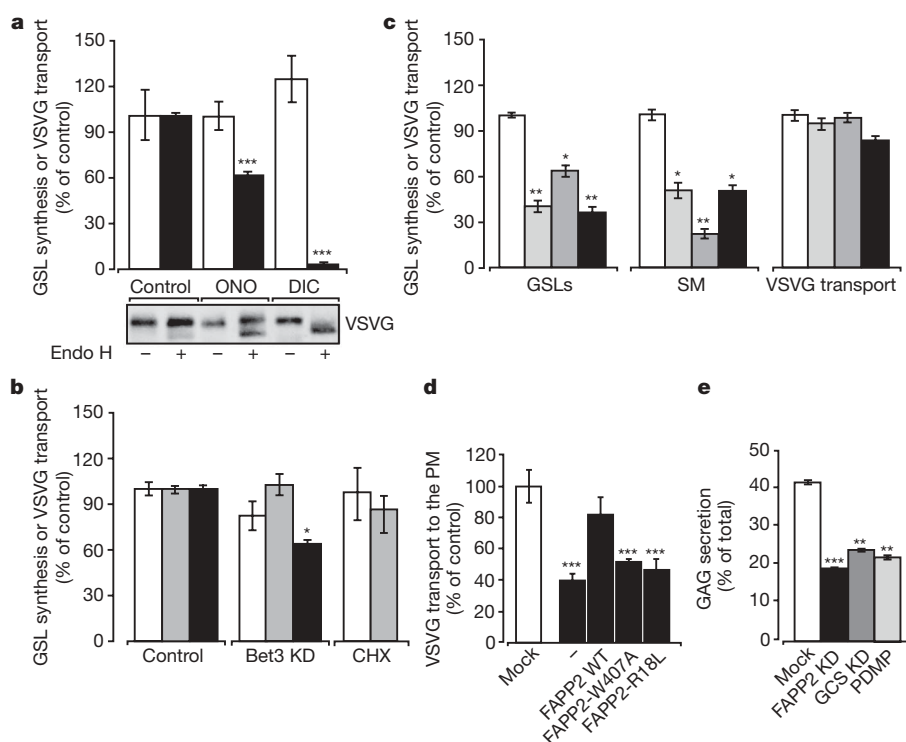


Figure 3 | GSL synthesis depends on PtdIns4P, and not on membrane trafficking at the GC. a–c, GSL synthesis: labelling with [³H]sphingosine for 60 min (a) or 120 min (b, c); VSVG transport: labelling with [³⁵S]methionine for 10 min, chase for 60 min; transport of VSVG quantified as fraction of VSVG acquiring resistance to endoglycosidase H (Endo H). **a,** Upper panel: cells mock-treated (control) or treated with 10 μ M ONO-RS-082 or 200 μ M dicoumarol 30 min before and during pulse–chase. Open columns, GSL synthesis; filled columns, VSVG transport. Lower panel: representative autoradiography of Endo-H-induced shift in VSVG mobility on SDS–PAGE. **b,** Cells were mock-treated (control), knocked down for Bet3 (Bet3 KD), or exposed to 50 μ g ml^{−1} cycloheximide (CHX) both 60 min before and during [³H]sphingosine pulse. Data are expressed as percentages of relative controls. White columns, GSL synthesis; grey columns, SM synthesis; black

columns, VSVG transport. **c,** Cells were mock-treated (white columns), knocked down for PI(4)KII α (light grey columns) or PI(4)KIII β (dark grey columns), or treated with 1 μ M PAO (black columns). **d, e,** GSL synthesis controls TGN-to-PM membrane trafficking. **d,** FAPP2 KD cells (filled columns) were transfected with siRNA-resistant forms of wild-type FAPP2 (FAPP2 WT) or FAPP2-W407A or FAPP2-R18L. Open column, mock-transfected cells. TGN-to-PM transport of VSVG was quantified by counting cells showing PM staining for VSVG. **e,** Cells were mock-treated, knocked down for FAPP2 or GCS, or treated with 75 μ M D-PDMP for 1 h. For release of [³⁵S]glycosaminoglycans, see Methods. Data are shown as means \pm s.d. for three independent experiments, each in duplicate ($n = 3$). Asterisks indicate statistical significance.

GSL synthesis was inhibited by GCS knockdown (Supplementary Fig. 3). The TGN-to-PM transport of glycosaminoglycans in GCS-knockdown cells was compared with control cells: it was impaired to a similar extent to that seen in FAPP2-knockdown cells and in cells treated with the GCS inhibitor phenyl-2-decanoyl-amino-3-morpholino-1-propanol-hydrochloride²⁷ (PDMP; Fig. 3e).

In agreement with our previous reports⁴, impairment of membrane trafficking by FAPP2 knockdown was through specific inhibition of TGN-to-PM transport, and did not involve other Golgi functions⁴ (Supplementary Fig. 5c, d). In addition, the ultrastructural organization of Golgi stacks was preserved in FAPP2-knockdown cells, with regular, although shorter, cisternae. However, the TGN was less developed and *trans* cisternae appeared swollen in FAPP2-knockdown cells (as well as in GCS-knockdown cells; Supplementary Fig. 7 and Supplementary Movies 1–3).

Taken together, our findings demonstrate that the GlcCer transfer activity of FAPP2 and GCS control TGN-to-PM transport, providing further evidence linking GSL metabolism to membrane trafficking^{27–32}.

Discussion

We show here that FAPP2, a protein binding PtdIns4P and ARF1 that controls TGN-to-PM trafficking, is an essential component of the

GSL synthetic machinery, because it mediates non-vesicular transport of the common GSL precursor GlcCer from its site of synthesis at the early Golgi to distal Golgi compartments, where most of the enzymes involved in complex GSLs assembly are located³³ (Fig. 4).

Our findings have many implications. The first is that newly synthesized proteins and GlcCer can use different transport pathways to cross the GC, with the former bound to move via membrane flow, and the latter, which is present in the cytosolic membrane leaflet, also amenable for cytosolic transport via lipid-transfer proteins (Fig. 4a). Together with the demonstration that SM synthesis requires the lipid-transfer protein CERT¹⁷ (previously known as GPBPA26 (ref. 34)), our data provide the most direct evidence that lipid-transfer proteins occupy key positions in the synthetic pathway of the sphingolipids. An issue that remains to be solved is how *cis*-to-*trans* vectoriality of the GlcCer transfer operated by FAPP2 is achieved. The enrichment of PtdIns4P in TGN membranes⁴ is likely to be important in this process.

Another important implication is that our data reveal a previously unrecognized and key component of the molecular machinery responsible for generation of the lipid identity of the PM. Indeed, although connected by intense bidirectional membrane trafficking, the PM–endosome–TGN and the early-Golgi–ER membrane systems have completely distinct lipid compositions, with the former highly enriched in GSLs, SM and cholesterol compared with the latter⁹.

The precursors of GSLs, SM and cholesterol itself, are the ligands for the family of lipid-binding proteins to which FAPP2 belongs, which includes CERT and oxysterol-binding protein 1 (OSBP1; ref. 35) (Fig. 5b). These members share a highly conserved PH domain that is responsible for their targeting to the TGN. CERT mediates the non-vesicular transport of ceramide from the ER to the TGN and is required for SM synthesis¹⁷, whereas OSBP1 is an oxysterol–cholesterol-binding protein with a proposed role in sensing and/or regulating cholesterol levels and distribution in cells. Being recruited to the TGN by PtdIns4P and ARF1, key regulators of anterograde transport from the TGN³⁶, FAPP2, CERT and OSBP1 can spatially and temporally coordinate and couple the synthesis of GSLs and SM, and possibly the delivery of cholesterol, to the anterograde transport of these lipids to the PM by means of membrane carriers (Fig. 5b).

METHODS SUMMARY

Lipid transfer assays *in vitro*. In the fluorescence resonance energy transfer assay, 10 µg of proteins, C12-BODIPY-lipids (1 mol%) and DiI-C18 (1,1'-dioctadecyl-3,3,3',3'-tetramethylindodicarbocyanine, perchlorate) were used to monitor the lipid-transfer activity in continuous real time, as described previously¹⁰. In the experiments using ARF1 and PtdIns4P, donor vesicles contained 2 mol% PtdIns, 3 mol% DiI-C18, 0.5 mol% BODIPY-GlcCer and 94.5 mol% POPC. Acceptor vesicles contained POPC or POPC with 2 mol% PtdIns4P or 0.2 µM ARF1 and 40 µM GTP. The lipid transfer assay with [¹⁴C]stearoyl-GlcCer and 1 mM proteins was performed as described previously³⁷.

GlcCer transfer assay in permeabilized cells. The C12-BODIPY-GlcCer delivery assay in streptolysin-O-permeabilized cells was performed as described in Methods. Protein loading with C-12-BODIPY-GlcCer from lipid films was performed as described previously³⁷.

Metabolic labelling with [³H]sphingosine and sphingolipid analysis. Metabolic labelling with [³H]sphingosine, GSL extraction and high-performance thin-layer chromatography (HPTLC) were performed as described³⁸. HPTLC plates were analysed by using a Raytest TLC Analyser V2.15, and then imaged with a Fujifilm BAS 1800-II phosphor imaging system. Endogenous sphingolipids were analysed by HPLC and LC-MS, as described^{39,40}.

Membrane trafficking assays. Transport of the temperature-sensitive mutant of VSVG (ts045-VSVG) and [³⁵S]glycosaminoglycan secretion were assessed as described previously^{41,42}. ts045-VSVG transport to the PM was evaluated by analysing the PM staining of at least 200 cells after 30 min of release from a 20 °C temperature block. Albumin secretion was assessed by immunoprecipitating radiolabelled albumin from medium or cell lysates of HepG2 cells pulsed with [³⁵S]methionine for 5 min and chased for 20 min, as described previously⁴³.

Immunolocalization of endogenous GlcCer. Ultrathin cell cryosectioning was performed at 4 °C, as described⁴⁴. Morphometric analysis was performed as described previously⁴⁵.

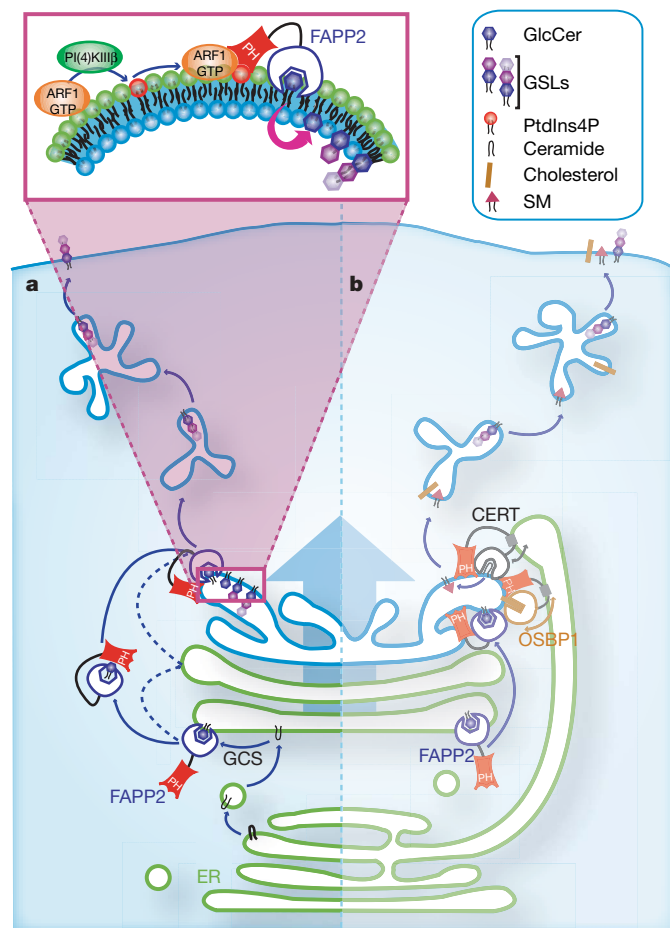


Figure 4 | Model of FAPP2 mechanism of action. Schematic diagram of the mechanism of action and regulation of FAPP2 (a) and other lipid-transfer proteins (b) at the GC. See the text and Supplementary Discussion for details. Inset: FAPP2 is recruited to the TGN by means of PH-domain binding to ARF1-GTP and PtdIns4P (produced by PI(4)Ks, including the ARF1-regulated PI(4)KIIIβ). Violet arrow, GlcCer translocator transferring GlcCer from the cytosolic to the luminal membrane leaflet; grey square in CERT and OSBP1, FFAT motif responsible for their targeting to the ER.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 June; accepted 20 July 2007.

Published online 8 August; corrected 6 September 2007.

1. Bard, F. & Malhotra, V. The formation of TGN-to-plasma-membrane transport carriers. *Annu. Rev. Cell Dev. Biol.* **22**, 439–455 (2006).
2. Luini, A., Ragnini-Wilson, A., Polishchuk, R. S. & De Matteis, M. A. Large pleiomorphic traffic intermediates in the secretory pathway. *Curr. Opin. Cell Biol.* **17**, 353–361 (2005).
3. De Matteis, M. A. & Godi, A. Protein–lipid interactions in membrane trafficking at the Golgi complex. *Biochim. Biophys. Acta* **1666**, 264–274 (2004).
4. Godi, A. *et al.* FAPPs control Golgi-to-cell-surface membrane traffic by binding to ARF and PtdIns(4)P. *Nature Cell Biol.* **6**, 393–404 (2004).
5. Vieira, O. V., Verkade, P., Manninen, A. & Simons, K. FAPP2 is involved in the transport of apical cargo in polarized MDCK cells. *J. Cell Biol.* **170**, 521–526 (2005).
6. Vieira, O. V. *et al.* FAPP2, cilium formation, and compartmentalization of the apical membrane in polarized Madin–Darby canine kidney (MDCK) cells. *Proc. Natl Acad. Sci. USA* **103**, 18556–18561 (2006).
7. Levine, T. P. & Munro, S. Targeting of Golgi-specific pleckstrin homology domains involves both PtdIns 4-kinase-dependent and -independent components. *Curr. Biol.* **12**, 695–704 (2002).
8. Brown, R. E. & Mattijus, P. Glycolipid transfer proteins. *Biochim. Biophys. Acta* **1771**, 746–760 (2007).
9. Holthuis, J. C., Pomorski, T., Raggars, R. J., Sprong, H. & Van Meer, G. The organizing potential of sphingolipids in intracellular membrane transport. *Physiol. Rev.* **81**, 1689–1723 (2001).
10. Nylund, M. *et al.* Molecular features of phospholipids that affect glycolipid transfer protein-mediated galactosylceramide transfer between vesicles. *Biochim. Biophys. Acta* **1758**, 807–812 (2006).
11. Wiedemann, C., Schafer, T. & Burger, M. M. Chromaffin granule-associated phosphatidylinositol 4-kinase activity is required for stimulated secretion. *EMBO J.* **15**, 2094–2101 (1996).
12. Brade, L., Vielhaber, G., Heinz, E. & Brade, H. *In vitro* characterization of anti-glucosylceramide rabbit antisera. *Glycobiology* **10**, 629–636 (2000).
13. Malinina, L., Malakhova, M. L., Teplov, A., Brown, R. E. & Patel, D. J. Structural basis for glycosphingolipid transfer specificity. *Nature* **430**, 1048–1053 (2004).
14. Klausner, R. D., Donaldson, J. G. & Lippincott-Schwartz, J. Brefeldin A: insights into the control of membrane traffic and organelle structure. *J. Cell Biol.* **116**, 1071–1080 (1992).
15. Boot, R. G. *et al.* Identification of the non-lysosomal glucosylceramidase as β -glucosidase 2. *J. Biol. Chem.* **282**, 1305–1312 (2007).
16. Yildiz, Y. *et al.* Mutation of β -glucosidase 2 causes glycolipid storage disease and impaired male fertility. *J. Clin. Invest.* **116**, 2985–2994 (2006).
17. Hanada, K. *et al.* Molecular machinery for non-vesicular trafficking of ceramide. *Nature* **426**, 803–809 (2003).
18. Mallard, F. *et al.* Direct pathway from early/recycling endosomes to the Golgi apparatus revealed through the study of Shiga toxin B-fragment transport. *J. Cell Biol.* **143**, 973–990 (1998).
19. Young, W. W. Jr, Lutz, M. S., Mills, S. E. & Lechler-Osborn, S. Use of brefeldin A to define sites of glycosphingolipid synthesis: GA2/GM2/GD2 synthase is trans to the brefeldin A block. *Proc. Natl Acad. Sci. USA* **87**, 6838–6842 (1990).
20. Mironov, A. A. *et al.* Dicumarol, an inhibitor of ADP-ribosylation of CtBP3/BARS, fragments golgi non-compact tubular zones and inhibits intra-golgi transport. *Eur. J. Cell Biol.* **83**, 263–279 (2004).
21. Brown, W. J., Chambers, K. & Doody, A. Phospholipase A₂ (PLA₂) enzymes in membrane trafficking: mediators of membrane shape and function. *Traffic* **4**, 214–221 (2003).
22. Drecktrah, D. & Brown, W. J. Phospholipase A₂ antagonists inhibit nocodazole-induced Golgi ministack formation: evidence of an ER intermediate and constitutive cycling. *Mol. Biol. Cell* **10**, 4021–4032 (1999).
23. Trucco, A. *et al.* Secretory traffic triggers the formation of tubular continuities across Golgi sub-compartments. *Nature Cell Biol.* **6**, 1071–1081 (2004).
24. Yu, S. *et al.* mBet3p is required for homotypic COPII vesicle tethering in mammalian cells. *J. Cell Biol.* **174**, 359–368 (2006).
25. Marra, P. *et al.* The biogenesis of the Golgi ribbon: the roles of membrane input from the ER and of GM130. *Mol. Biol. Cell* **18**, 1595–1608 (2007).
26. Toth, B. *et al.* Phosphatidylinositol 4-kinase III β regulates the transport of ceramide between the endoplasmic reticulum and Golgi. *J. Biol. Chem.* **281**, 36369–36377 (2006).
27. Rosenwald, A. G., Machamer, C. E. & Pagano, R. E. Effects of a sphingolipid synthesis inhibitor on membrane transport through the secretory pathway. *Biochemistry* **31**, 3581–3590 (1992).
28. Sprong, H. *et al.* Glycosphingolipids are required for sorting melanosomal proteins in the Golgi complex. *J. Cell Biol.* **155**, 369–380 (2001).
29. Schwarz, A. & Futerman, A. H. Distinct roles for ceramide and glucosylceramide at different stages of neuronal growth. *J. Neurosci.* **17**, 2929–2938 (1997).
30. Boldin, S. A. & Futerman, A. H. Up-regulation of glucosylceramide synthesis upon stimulation of axonal growth by basic fibroblast growth factor. Evidence for post-translational modification of glucosylceramide synthase. *J. Biol. Chem.* **275**, 9905–9909 (2000).
31. Chang, M. C., Wisco, D., Ewers, H., Norden, C. & Winckler, B. Inhibition of sphingolipid synthesis affects kinetics but not fidelity of L1/NG2CAM transport along direct but not transcytotic axonal pathways. *Mol. Cell. Neurosci.* **31**, 525–538 (2006).
32. Tamboli, I. Y. *et al.* Inhibition of glycosphingolipid biosynthesis reduces secretion of the β -amyloid precursor protein and amyloid β -peptide. *J. Biol. Chem.* **280**, 28110–28117 (2005).
33. Lannert, H., Gorgas, K., Meissner, I., Wieland, F. T. & Jeckel, D. Functional organization of the Golgi apparatus in glycosphingolipid biosynthesis. Lactosylceramide and subsequent glycosphingolipids are formed in the lumen of the late Golgi. *J. Biol. Chem.* **273**, 2939–2946 (1998).
34. Raya, A. *et al.* Goodpasture antigen-binding protein, the kinase that phosphorylates the Goodpasture antigen, is an alternatively spliced variant implicated in autoimmune pathogenesis. *J. Biol. Chem.* **275**, 40392–40399 (2000).
35. De Matteis, M. A., Di Campli, A. & D'Angelo, G. Lipid-transfer proteins in membrane trafficking at the Golgi complex. *Biochim. Biophys. Acta* **1771**, 761–768 (2007).
36. De Matteis, M. A. & Godi, A. PI-loting membrane traffic. *Nature Cell Biol.* **6**, 487–492 (2004).
37. Malakhova, M. L. *et al.* Point mutational analysis of the liganding site in human glycolipid transfer protein. Functionality of the complex. *J. Biol. Chem.* **280**, 26312–26320 (2005).
38. Sala, G., Dupre, T., Seta, N., Codogno, P. & Ghidoni, R. Increased biosynthesis of glycosphingolipids in congenital disorder of glycosylation Ia (CDG-Ia) fibroblasts. *Pediatr. Res.* **52**, 645–651 (2002).
39. Wing, D. R. *et al.* High-performance liquid chromatography analysis of ganglioside carbohydrates at the picomole level after ceramide glycanase digestion and fluorescent labeling with 2-aminobenzamide. *Anal. Biochem.* **298**, 207–217 (2001).
40. Bielawski, J., Szulc, Z. M., Hannun, Y. A. & Bielawska, A. Simultaneous quantitative analysis of bioactive sphingolipids by high-performance liquid chromatography–tandem mass spectrometry. *Methods* **39**, 82–91 (2006).
41. Godi, A. *et al.* ADP ribosylation factor regulates spectrin binding to the Golgi complex. *Proc. Natl Acad. Sci. USA* **95**, 8607–8612 (1998).
42. Marra, P. *et al.* The GM130 and GRASP65 Golgi proteins cycle through and define a subdomain of the intermediate compartment. *Nature Cell Biol.* **3**, 1101–1113 (2001).
43. Lodish, H. F. & Kong, N. Glucose removal from N-linked oligosaccharides is required for efficient maturation of certain secretory glycoproteins from the rough endoplasmic reticulum to the Golgi complex. *J. Cell Biol.* **98**, 1720–1729 (1984).
44. Watt, S. A., Kular, G., Fleming, I. N., Downes, C. P. & Lucocq, J. M. Subcellular localization of phosphatidylinositol 4,5-bisphosphate using the pleckstrin homology domain of phospholipase C δ 1. *Biochem. J.* **363**, 657–666 (2002).
45. Polishchuk, E. V., Di Pentima, A., Luini, A. & Polishchuk, R. S. Mechanism of constitutive export from the golgi: bulk flow via the formation, protrusion, and en bloc cleavage of large trans-golgi network tubular domains. *Mol. Biol. Cell* **14**, 4470–4485 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Luini, R. Ghidoni, P. Viani, L. Riboni for discussions; A. Luini, D. Corda and M. Gimona for critical reading of the manuscript; T. Scanu for sharing the data on Bet3 knockdown; C. Iurisci for technical assistance; A. Spaar for the bioinformatic analysis; G. Perinetti for the statistic analysis of the data; J. P. Slotte for the fluorescence facility; J. Saus for anti-CERT antibody; C. P. Berrie for editorial assistance; and E. Fontana for the artwork. This work was supported by the Telethon Electron Microscopy Core Facility, by a fellowship from FIRC to G.D.A., and by Telethon, AIRC, the Academy of Finland, Sigrid Jusélius, Magnus Ehrnrooth, K. Albin Johansson Foundations, Medicinska Undersökningsföreningen Liv och Hälsa, ISB Graduate School and Åbo Akademi University.

Author Contributions M.A.D.M. supervised the entire project and wrote the manuscript with G.D.A. and with comments from all coauthors; G.D.A. designed and conducted the experiments of sphingolipid labelling; G.D.A. designed the experiments of membrane trafficking, which were conducted by A.D.C. and G.D.T.; G.D.A. designed the cloning of FAPP2, which was conducted by G.D.T. and M.S.; G.D.T. and M.S. prepared all the constructs, recombinant proteins and anti-FAPP1, FAPP2, PI(4)KIII β , PI(4)KII α , BET3 and GM130 antibodies; E.P. and R.P. designed and conducted the experiments for electron microscopy; A.G. performed the trafficking experiment in FAPP1-knockdown cells; P.M. designed the experiments of intervesicular lipid transfer, which were conducted by G.W.; A.S. and C.C.C. conducted the HPLC measurements of GSLs under the supervision of F.M.P.; J.B. conducted the sphingolipid analysis by LC–MS under the supervision of Y.A.H.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.D.M. (dematteis@negrisud.it).

METHODS

Materials. All chemical reagents were of analytical grade or higher and purchased from Sigma unless otherwise specified. Cell culture media were from Invitrogen. Polyclonal antibodies against human FAPP1, FAPP2, PI(4)KIII β , PI(4)KII α , Bet3 and GM130 were raised in rabbits, with glutathione S-transferase fusion proteins as immunogens. All were affinity purified on their corresponding immunogens. The rabbit polyclonal antibody against GlcCer was from Glycobiotech GmbH. Rabbit polyclonal antibodies against GCS were from Exalpha Biologicals. The anti-ts045-VSVG clone P5D4, Cy3-conjugated and anti-Flag M5 monoclonal antibodies, and the anti-rabbit and anti-mouse IgG Cy3-conjugated antibodies were from Sigma. Sheep polyclonal antibodies against TGN46 were from AbD Serotech. The mouse monoclonal antibody against CERT was a gift from J. Saus. The Alexa 488 goat anti-mouse and anti-rabbit IgG (H+L) antibodies were from Molecular Probes. All unlabelled purified lipids were from Avanti Polar Lipids. BODIPY-labelled lipids were from Invitrogen Molecular Probes. [14 C]Stearoyl-GlcCer was from American Radiolabelled Chemicals, and [3 H]sphingosine was from PerkinElmer.

Binding assay for dichlorotriazinylaminofluorescein-conjugated StxB. HeLa cells that were mock-treated or siRNA-treated for the knockdown of FAPP2, GCS or CERT were incubated on ice for 25 min with 0.5 μ g ml $^{-1}$ dichlorotriazinylaminofluorescein-conjugated StxB. The cells were washed three times with ice-cold PBS to remove the unbound toxin and were subsequently fixed in 4% paraformaldehyde for 10 min (ref. 18). The samples were then processed for immunofluorescence and analysed with an LSM 510 confocal microscope (Zeiss).

Electron microscopy, immunoelectron microscopy, electron tomography and morphometric analysis. Detection of green fluorescent protein-conjugated FAPP2 by immunoelectron microscopy was performed in transfected HeLa cells as described previously⁴⁵. To distinguish membranes of endosomal origin from the TGN in the *trans*-Golgi area, cells were incubated with 25 μ g ml $^{-1}$ transferrin

conjugated with horseradish peroxidase, as described⁴⁶. The surface areas of round and tubular profiles at the TGN and of corresponding Golgi stacks were measured and compared in 30 control and 30 FAPP2-knockdown cells. Electron tomography of chemically fixed samples of HeLa cells was performed on sections 250 nm thick, as described previously²³.

RNA interference. The siRNAs for human FAPP2 (AF380162), CERT (NM_031361), GCS (NM_003358), Bet3 (NM_014408), PI(4)KIII β (NM_002651), PI(4)KII α (NM_018425.2) and FAPP1 (NM_019091) comprised mixtures of four siRNA duplexes (Supplementary Fig. 3) and were obtained from Dharmacon. HeLa and COS7 cells were plated at 30% confluence in 12-well plates and transfected with 120–150 pmol of siRNAs with Oligofectamine (Invitrogen), in accordance with the manufacturer's protocol. At 72 h after the initial treatment with siRNA, the cells were processed directly for the different experiments.

To test the effects of FAPP2-siRNA-resistant forms, a single siRNA duplex was used to knock down endogenous FAPP2. The duplex sequence was 5'-GAAUUGAUGUGGGAACUUU-3'. The 3 \times Flag-FAPP2-519 and the single-point mutant W407A of 3 \times Flag-FAPP2-519 were mutagenized to produce mismatches in the duplex region without inducing any amino-acid change in the protein sequence. The DNA sequence in the duplex region after mutations was (mutated bases in bold): 5'-GAAATCGATGTCGAACCTT-3'. At 48 h after the initial siRNA treatment, the cells were transfected with plasmids expressing 3 \times Flag-FAPP2 or the siRNA-resistant versions of FAPP2, and the levels of the transfected and endogenous FAPP2 were assessed by western blotting with specific antibodies (Supplementary Fig. 3b).

Statistical analysis. Two-tailed Student *t*-tests were applied to the data. Asterisk, $P \leq 0.05$; two asterisks, $P \leq 0.01$; three asterisks, $P \leq 0.005$.

46. Millar, C. A. *et al.* Adipsin and the glucose transporter GLUT4 traffic to the cell surface via independent pathways in adipocytes. *Traffic* 1, 141–151 (2000).

LETTERS

Entanglement of single-atom quantum bits at a distance

D. L. Moehring¹, P. Maunz¹, S. Olmschenk¹, K. C. Younge¹, D. N. Matsukevich¹, L.-M. Duan¹ & C. Monroe^{1,2}

Quantum information science involves the storage, manipulation and communication of information encoded in quantum systems, where the phenomena of superposition and entanglement can provide enhancements over what is possible classically^{1,2}. Large-scale quantum information processors require stable and addressable quantum memories, usually in the form of fixed quantum bits (qubits), and a means of transferring and entangling the quantum information between memories that may be separated by macroscopic or even geographic distances. Atomic systems are excellent quantum memories, because appropriate internal electronic states can coherently store qubits over very long timescales. Photons, on the other hand, are the natural platform for the distribution of quantum information between remote qubits, given their ability to traverse large distances with little perturbation. Recently, there has been considerable progress in coupling small samples of atomic gases through photonic channels^{2,3}, including the entanglement between light and atoms^{4,5} and the observation of entanglement signatures between remotely located atomic ensembles^{6–8}. In contrast to atomic ensembles, single-atom quantum memories allow the implementation of conditional quantum gates through photonic channels^{2,9}, a key requirement for quantum computing. Along these lines, individual atoms have been coupled to photons in cavities^{2,10–12}, and trapped atoms have been linked to emitted photons in free space^{13–17}. Here we demonstrate the entanglement of two fixed single-atom quantum memories

separated by one metre. Two remotely located trapped atomic ions each emit a single photon, and the interference and detection of these photons signals the entanglement of the atomic qubits. We characterize the entangled pair by directly measuring qubit correlations with near-perfect detection efficiency. Although this entanglement method is probabilistic, it is still in principle useful for subsequent quantum operations and scalable quantum information applications^{18–20}.

In each of two congeneric radio-frequency ion traps, we trap and laser-cool a single $^{171}\text{Yb}^+$ ion²¹. Each ion is cooled to near the Doppler limit via laser light at 369.5 nm tuned just redward of the $^2\text{S}_{1/2} \leftrightarrow ^2\text{P}_{1/2}$ atomic resonance. The $^2\text{P}_{1/2}$ level also has a decay channel to the $^2\text{D}_{3/2}$ state with a branching ratio of ~ 0.005 (ref. 21). When this decay occurs, the ion is pumped back to the $^2\text{S}_{1/2}$ level via the application of a 935.2 nm laser, as depicted in Fig. 1a. A magnetic field of $B \approx 5.5$ G provides a quantization axis for definition of the photon polarization and the internal atomic qubit levels, stored in the hyperfine levels of the $^2\text{S}_{1/2}$ ground state. This magnetic field also suppresses coherent dark state trapping in the $^2\text{S}_{1/2}$ levels during Doppler cooling and atomic state detection²².

Figure 2 shows a diagram of the relevant energy levels of $^{171}\text{Yb}^+$ along with the step-by-step description of the experimental procedure. After Doppler cooling, pulses of light tuned to the $^2\text{S}_{1/2} \leftrightarrow ^2\text{P}_{1/2}$ transitions initialize, excite and detect the internal states of the ions. First, a 500 ns pulse of light resonant with the $^2\text{S}_{1/2} |F=1\rangle \leftrightarrow ^2\text{P}_{1/2}$

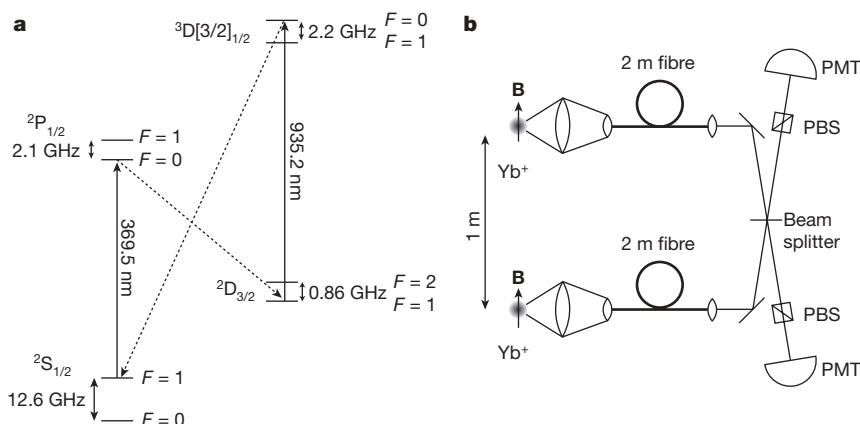


Figure 1 | Experimental apparatus. **a**, Relevant energy levels for $^{171}\text{Yb}^+$. The $^2\text{S}_{1/2} \leftrightarrow ^2\text{P}_{1/2}$ transition is driven by light at 369.5 nm. A frequency-doubled continuous-wave diode laser is used for Doppler cooling, state initialization, and state detection of the ion, whereas the excitation of the ion for single photon generation is accomplished with a mode-locked, frequency-doubled Ti:sapphire laser with a pulse duration of 2 ps (ref. 23). When excited to $^2\text{P}_{1/2}$, the ion can decay to the $^2\text{D}_{3/2}$ level with a branching ratio of ~ 0.005 (ref. 21). A diode laser at 935.2 nm pumps the ion out of this state through the $^3\text{D}[3/2]_{1/2}$ level²¹. **b**, Two ions are trapped in independent vacuum chambers separated by approximately 1 m. Spontaneously emitted photons

from each ion are collected by an $f/2.1$ imaging lens and coupled into single-mode fibres. The polarization of each emitted photon is defined with respect to the applied magnetic field B oriented perpendicularly to the collection direction. Polarization is maintained through the fibres and can be adjusted via polarization-controlling paddles. The output of each fibre is spatially mode-matched on a 50/50 non-polarizing beam splitter leading to an interference contrast of greater than 97%. Polarizing beam splitters (PBS) are used to filter out the photons of unwanted polarization and the remaining photons are detected on photon-counting PMTs.

¹FOCUS Center and Department of Physics, University of Michigan, Ann Arbor, Michigan 48109-1040, USA. ²JQI and Department of Physics, University of Maryland, College Park, Maryland 20742, USA.

$|F=1\rangle$ transition prepares each ion in the state $|F, m_F\rangle = |0,0\rangle$ with more than 99% efficiency²¹. Here F is the total angular momentum and m_F is its projection along the quantization axis. Next, a 2 ps σ^- -polarized laser pulse from a frequency-doubled, mode-locked Ti-sapphire laser excites the ion to the $^2P_{1/2} |1, -1\rangle$ state on a timescale much shorter than the excited state lifetime of $\tau = 8.1$ ns. The ion is prepared in the excited state with an excitation probability of $P_{\text{exc}} \approx 0.5$ (see Methods), and spontaneously decays to either the $^2S_{1/2} |1, -1\rangle$ state while emitting a π -polarized photon or to the $|1,0\rangle \equiv |\uparrow\rangle$ and $|0,0\rangle \equiv |\downarrow\rangle$ states while emitting a σ^- -polarized photon (Fig. 2b). The spontaneously emitted photon at 369.5 nm is collected with an $f/2.1$ imaging lens along a direction perpendicular to the quantization axis and is coupled into a single-mode fibre. Along this direction, the polarizations of the π and σ^- decay channels are orthogonal¹³, and the π -polarized photons are filtered out with polarizers. The resulting entangled state between the ion and the photon is $(|\uparrow\rangle|\nu_\uparrow\rangle - |\downarrow\rangle|\nu_\downarrow\rangle)/\sqrt{2}$, where $|\nu_\downarrow\rangle$ and $|\nu_\uparrow\rangle$ refer to the two resolved frequencies comprising the photonic qubit⁹. The negative sign is a result of the Clebsch–Gordon coefficients, and the two frequency components are separated by the 12.6 GHz ground-state hyperfine splitting.

The output modes of the fibres from each trap are directed onto the two input ports of a 50/50 non-polarizing beam splitter with a transverse spatial mode overlap leading to an interference contrast of more than 97% (Fig. 1b)²³. The photons emerging from the beam splitter are sent through polarizers to filter out the π -polarized decay channel and then to photon-counting photomultiplier tubes (PMTs), each of quantum efficiency $\eta \approx 0.15$.

While the end-to-end coupling efficiency through each fibre is only $\zeta \approx 0.2$ (including the ~ 0.1 dB m^{-1} fibre attenuation), the use of single-mode fibres is essential for the rejection of photons in other spatial modes that would not properly interfere and thereby lower the entanglement fidelity^{15,23}. Temporal mode-matching of photons is accomplished by matching the arrival times of the photon from each ion on the beam splitter to better than 30 ps. Compared to the 8.1 ns photon duration, the remaining temporal mismatch corresponds to a decrease in the mode overlap of under 1%. Finally, spectral matching of the photonic qubits is accomplished by equalizing the magnetic field at the traps to better than 30 mG, resulting in a photonic frequency mismatch of less than 0.2% of the $1/(2\pi\tau) \approx 20$ MHz photon bandwidth. Doppler broadening of the photon emission, from both residual motion of the Doppler-cooled ions and micromotion²⁴, is expected to affect the interference by well under 1%.

When each ion emits a photon into the mode of interest, the quantum state of the system before the photons interact on the

50/50 beam splitter is:

$$\frac{1}{2} [(|\uparrow\rangle_a |\nu_\uparrow\rangle_a - |\downarrow\rangle_a |\nu_\downarrow\rangle_a) \otimes (|\uparrow\rangle_b |\nu_\uparrow\rangle_b - |\downarrow\rangle_b |\nu_\downarrow\rangle_b)] = \frac{1}{2} (|\Phi^+\rangle_{\text{atom}} |\Phi^+\rangle_{\text{photon}} + |\Phi^-\rangle_{\text{atom}} |\Phi^-\rangle_{\text{photon}} - |\Psi^+\rangle_{\text{atom}} |\Psi^+\rangle_{\text{photon}} - |\Psi^-\rangle_{\text{atom}} |\Psi^-\rangle_{\text{photon}}) \quad (1)$$

where $|\Phi^\pm\rangle_{\text{atom}} = (|\uparrow\rangle_a |\uparrow\rangle_b \pm |\downarrow\rangle_a |\downarrow\rangle_b)/\sqrt{2}$ and $|\Psi^\pm\rangle_{\text{atom}} = (|\uparrow\rangle_a |\downarrow\rangle_b \pm |\downarrow\rangle_a |\uparrow\rangle_b)/\sqrt{2}$ are the maximally entangled Bell states for the ions, with corresponding definitions for the photons. With the photon modes matched on the 50/50 beam splitter, the photons exit on different ports only if they are in the antisymmetric state $|\Psi^-\rangle_{\text{photon}} = (|\nu_\uparrow\rangle_a |\nu_\downarrow\rangle_b - |\nu_\downarrow\rangle_a |\nu_\uparrow\rangle_b)/\sqrt{2}$, respecting the symmetry of the overall photonic wavefunction²⁵. Therefore, coincident photon detection in the two output ports of this beam splitter ideally projects the ions onto $|\Psi^-\rangle_{\text{atom}}$ (ref. 26). (This result assumes equal path lengths from each ion to the beam splitter. A simple extension considering differing path lengths can be found in the Methods.) We note that it is the absence of interference between the two different frequency components of each photon that allows for simultaneous detection of a photon on both detectors, because two photons of the same polarization and frequency cannot emerge from the beam splitter along separate paths^{23,27}. Following a heralded entanglement event, near-resonant microwave pulses coherently rotate the trapped ion qubits and prepare them for measurement in different bases. The atomic qubit measurement is performed using standard trapped ion fluorescence techniques with detection efficiency greater than 97% (ref. 21) (Fig. 2c).

We first measure the state of the two ions after the coincident photon detection without microwave rotations. The expected resulting ion–ion entangled state is $|\Psi^-\rangle_{\text{atom}}$, so the atomic wavefunction should have odd parity ($|\uparrow\rangle_a |\downarrow\rangle_b$ or $|\downarrow\rangle_a |\uparrow\rangle_b$). The probability distribution from 274 coincidence detection events is shown in Fig. 3 with a resulting probability of measuring odd parity $p_{\downarrow\uparrow} + p_{\uparrow\downarrow} = 0.78 \pm 0.02$. Here, p_{ab} refers to the probability of measuring the two atomic qubits (ions a and b) in the given spin states.

To verify the entanglement, we repeat the experiment and measure in a rotated basis. Each ion is rotated through a Bloch polar angle of $\theta = \pi/2$ by applying microwave pulses of duration ~ 4 μs near the measured $|\uparrow\rangle_i \leftrightarrow |\downarrow\rangle_i$ frequency splitting of 12.642821 GHz (Fig. 2c). The two atomic qubit transition frequencies are matched to better than 100 Hz and through the use of magnetic-field-insensitive ‘clock states’ as the qubit, the transition frequencies are essentially static over the course of the experiment^{21,28–30}. We vary the relative phase $\Delta\phi = \phi_b - \phi_a$ of the rotations at the two ions by detuning the applied microwaves by 10 kHz from resonance and delaying the microwave pulse on one ion with respect to the other. Here ϕ_i refers to the

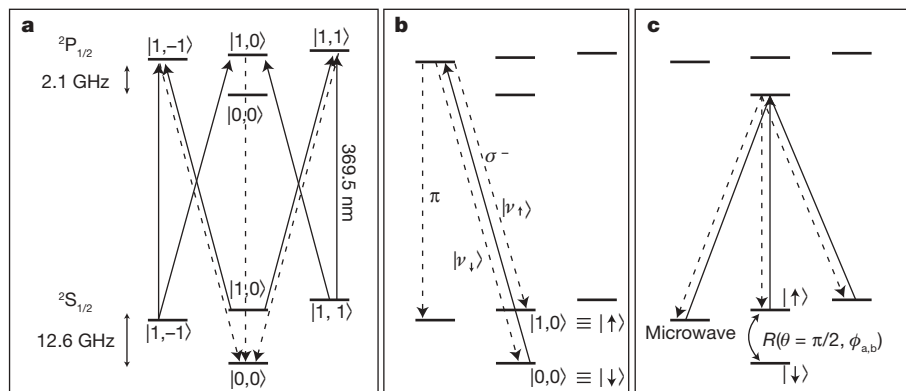


Figure 2 | Experimental procedure. **a**, Each atomic qubit is initialized to the $|F, m_F\rangle = |0,0\rangle$ hyperfine ground state via a 500 ns optical pumping pulse resonant with the $^2S_{1/2} |F=1\rangle \leftrightarrow ^2P_{1/2} |F=1\rangle$ transition including all polarizations. **b**, Each ion is excited with a 2 ps σ^- -polarized optical pulse, resulting in the spontaneous emission of at most a single photon either to the $^2S_{1/2} |1, -1\rangle$ state while emitting a π -polarized photon or to the $|\uparrow\rangle$ and $|\downarrow\rangle$ states while emitting a σ^- -polarized photon. The π -polarized photon is

filtered out by a polarizer resulting in the entangled state $(|\uparrow\rangle|\nu_\uparrow\rangle - |\downarrow\rangle|\nu_\downarrow\rangle)/\sqrt{2}$. The steps in **a** and in **b** are repeated on both ions until simultaneous detection occurs on two PMTs. **c**, Upon simultaneous detection of a photon on each PMT, an optional microwave rotation pulse prepares each atomic qubit for measurement in different bases, followed by measurement of the atomic qubits using standard trapped ion fluorescence techniques, thus resonantly driving all $^2S_{1/2} |F=1\rangle \leftrightarrow ^2P_{1/2} |F=0\rangle$ transitions.

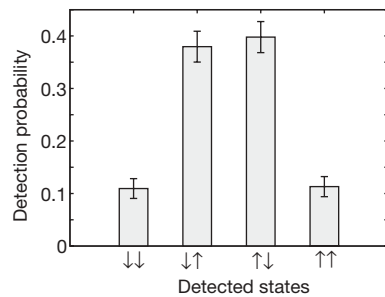


Figure 3 | Unrotated basis results. Measured probabilities in the unrotated basis (no atomic qubit rotation before measurement) conditioned upon coincident detection of photons on each PMT (photon pairs with less than 16 ns detection-time difference). The measured probabilities are $p_{\downarrow\downarrow} = 0.11 \pm 0.02$, $p_{\downarrow\uparrow} = 0.38 \pm 0.03$, $p_{\uparrow\downarrow} = 0.40 \pm 0.03$ and $p_{\uparrow\uparrow} = 0.11 \pm 0.02$, so odd parity is found with a probability of $p_{\downarrow\uparrow} + p_{\uparrow\downarrow} = 0.78 \pm 0.02$. The errors are statistical (standard error) and are collected from 274 coincidence detection events.

microwave phase on ion $i = a, b$ and a 100 μs delay results in a phase difference of $\Delta\phi = 2\pi$. Figure 4 displays the resulting oscillations of the measured atomic qubit parity as a function of the relative phase of the $\pi/2$ rotations. The contrast of this oscillation is directly related to the coherence between $|\downarrow_a\uparrow_b\rangle$ and $|\uparrow_a\downarrow_b\rangle$.

From these measurements, we calculate a fidelity of $\mathcal{F} = 0.63 \pm 0.03$ and a lower bound on the entanglement of formation to be $\mathcal{E} \geq 0.12 \pm 0.03$, as described in the Methods. The results are limited mainly by dark counts on the PMTs that lead to false events in $\sim 20\%$ of the measured coincidence events. Other factors contributing to the decrease in fidelity include atomic state detection errors ($< 3\%$ for each ion), imperfect mode-matching on the 50/50 beam splitter (3%), mixing of the photon polarizations owing to the non-zero solid angle (1.5% of detected photons result from a $\Delta m_F = 0$ decay), excitations to the wrong atomic state ($\sim 1\%$), and imperfect rotations of the atomic qubit ($\sim 1\%$). Sources of error from imperfections in the optical fibres and magnetic field fluctuations are estimated to affect the measured entanglement by less than 1%.

As mentioned above, the remote-atom entanglement is a probabilistic process. The success probability P in a given trial depends on the efficiency of generating a single photon from each ion in a specific mode and detecting the photons in coincidence. In our excitation scheme (Fig. 2b), each ion has a probability $P_{\text{exc}} \approx 0.5$ of emitting a single photon after the excitation pulse, and $\rho \approx 0.995$ of the emitted photons are at 369.5 nm due to the branching ratio to the $^2D_{3/2}$ state. The detection probability of each photon is given by the light collection solid angle $\Delta\Omega/4\pi \approx 0.02$, coupling efficiency and transmission through the single-mode fibre ζ , transmission through other optical elements $T \approx 0.8$, and the quantum efficiency of the detectors η . In addition, half of the collected photons are π -polarized and are filtered out by the polarizer¹³. Finally, because only the $|\Psi^-\rangle_{\text{photon}}$ state results in the two photons exiting the 50/50 beam splitter in different output ports, there is an additional factor of 1/4 in our success probability: $P = (1/4)[(1/2)\eta\zeta T\rho P_{\text{exc}}(\Delta\Omega/4\pi)]^2 \approx (0.25)[(0.5)(0.15)(0.2)(0.8)(0.995)(0.5)(0.02)]^2 \approx 3.6 \times 10^{-9}$. With an experiment repetition rate of $R \approx 5.5 \times 10^5 \text{ s}^{-1}$, this results in a heralded entanglement event approximately every 8.5 min.

This rate is proportional to the square of the probability of measuring an emitted photon, so improvements in the generation of single photons in the desired mode can significantly increase this yield. Possible improvements include increasing the probability of excitation to unity by using an alternative excitation scheme or collecting the emitted photons along the quantization axis where the radiation strength of the emitted photons is greatest. However, the most substantial improvement would be to place the ion within an optical cavity, which would allow the effective solid angle $\Delta\Omega/4\pi$ to approach unity. Not only would these changes increase the success

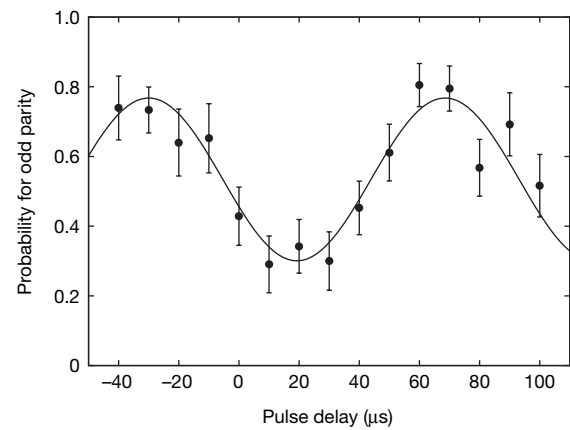


Figure 4 | Rotated bases results. Probability of measuring odd parity of the atomic qubits after rotations on each ion by a polar angle of $\pi/2$ on the Bloch sphere. The horizontal axis corresponds to the delay of the microwave pulse on one ion with respect to the other. Because of the 10 kHz detuning of the applied microwave radiation from resonance, the relative phase of the microwave pulses, $\Delta\phi = \phi_b - \phi_a$, has a period of 100 μs . The solid line is a fit to the data resulting in a contrast of 0.47 ± 0.05 , where the absolute phase of the interference pattern depends on the difference in microwave transmission lines to each ion. Error bars are statistical (standard error) and the results are from a total of 502 coincidence detection events.

probability, they would also substantially improve the degree of entanglement by lowering the relative contribution of dark count events.

One possible use of this entanglement scheme is a loophole-free Bell inequality violation, with matter qubits prepared in remote locations and measured with high efficiency²⁶. However, successful demonstration of such a violation would require higher entanglement fidelities over much larger distances than that reported here. Additionally, this heralded entanglement could form the basis of a quantum repeater, networking intermediately located ions^{2,18}. Owing to the long coherence and storage times of trapped atomic ions, with the improvements listed above, such a network of remotely entangled ions could be used for scalable quantum computation and communication^{9,18,19}.

METHODS SUMMARY

The use of an ultrafast laser for excitation of the ion to the $^2P_{1/2} |1, -1\rangle$ state is necessary to ensure each ion scatters at most one photon per excitation pulse²³. In our excitation scheme, the broad bandwidth of the ultrafast laser leads to a maximum excitation probability of $P_{\text{exc}} = 50\%$, where the remaining 50% is in a coherent superposition of $|\uparrow\rangle$ and $|\downarrow\rangle$, and does not result in an emitted photon. While P_{exc} can in principle be increased to unity by using alternate atomic states or transitions, the above excitation scheme is sufficient for the creation of photon pairs, and thus for heralded entanglement. The resulting entanglement fidelity and entanglement of formation are calculated from the elements of the density matrix obtained by measuring the ions in the rotated and unrotated bases after the heralded entanglement. The form of the entangled state as given in the text may require an additional phase factor if the path lengths from each ion to the beam splitter are different. In the most general case, coincident photon detection projects the ions onto $(e^{-i\Delta k \Delta x} |\downarrow_a\uparrow_b\rangle - |\uparrow_a\downarrow_b\rangle)/\sqrt{2}$, where Δk is the difference in wavevectors of the two photon-frequency qubits and Δx is the difference in photon path length from each ion to the beam splitter.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 May; accepted 14 July 2007.

- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, UK, 2000).
- Zoller, P. *et al.* Quantum information processing and communication. *Eur. Phys. J. D* **36**, 203–228 (2005).
- Duan, L.-M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).

4. Sherson, J., Julsgaard, B. & Polzik, E. S. Deterministic atom-light quantum interface. *Adv. At. Mol. Opt. Phys.* **54**, 82–130 (2006).
5. Jenkins, S. D. *et al.* Quantum telecommunication with atomic ensembles. *J. Opt. Soc. Am. B* **24**, 316–323 (2007).
6. Julsgaard, B., Kozhekin, A. & Polzik, E. S. Experimental long-lived entanglement of two macroscopic objects. *Nature* **413**, 400–403 (2001).
7. Chou, C. W. *et al.* Measurement-induced entanglement for excitation stored in remote atomic ensembles. *Nature* **438**, 828–832 (2005).
8. Matsukevich, D. N. *et al.* Entanglement of remote atomic qubits. *Phys. Rev. Lett.* **96**, 030405 (2006).
9. Duan, L.-M. *et al.* Probabilistic quantum gates between remote atoms through interference of optical frequency qubits. *Phys. Rev. A* **76**, 062324 (2006).
10. Berman, P. (ed.) *Cavity Quantum Electrodynamics* (Academic Press, San Diego, California, 1994).
11. McKeever, J. *et al.* Deterministic generation of single photons from one atom trapped in a cavity. *Science* **303**, 1992–1994 (2004).
12. Wilk, T., Webster, S. C., Kuhn, A. & Rempe, G. Single-Atom Single-Photon Quantum Interface. *Science* **317**, 488–490 (2007).
13. Blinov, B. B., Moehring, D. L., Duan, L.-M. & Monroe, C. Observation of entanglement between a single trapped atom and a single photon. *Nature* **428**, 153–157 (2004).
14. Moehring, D. L., Madsen, M. J., Blinov, B. B. & Monroe, C. Experimental bell inequality violation with an atom and a photon. *Phys. Rev. Lett.* **93**, 090410 (2004).
15. Beugnon, J. *et al.* Quantum interference between two single photons emitted by independently trapped atoms. *Nature* **440**, 779–782 (2006).
16. Volz, J. *et al.* Observation of entanglement of a single photon with a trapped atom. *Phys. Rev. Lett.* **96**, 030404 (2006).
17. Moehring, D. L. *et al.* Quantum networking with photons and trapped atoms. *J. Opt. Soc. Am. B* **24**, 300–315 (2007).
18. Duan, L.-M., Blinov, B. B., Moehring, D. L. & Monroe, C. Scaling trapped ions for quantum computation with probabilistic ion-photon mapping. *Quant. Inf. Comp.* **4**, 165–173 (2004).
19. Duan, L.-M. & Raussendorf, R. Efficient quantum computation with probabilistic quantum gates. *Phys. Rev. Lett.* **95**, 080503 (2005).
20. Barrett, S. D. & Kok, P. Efficient high-fidelity quantum computation using matter qubits and linear optics. *Phys. Rev. A* **71**, 060310(R) (2005).
21. Olmschenk, S. *et al.* Manipulation and detection of a trapped Yb^+ ion hyperfine qubit. Preprint at (<http://arxiv.org/abs/0708.0657>) (2007).
22. Berkeland, D. J. & Boshier, M. G. Destabilization of dark states and optical spectroscopy in Zeeman-degenerate atomic systems. *Phys. Rev. A* **65**, 033413 (2002).
23. Maunz, P., Moehring, D. L., Olmschenk, S., Younge, K. C., Matsukevich, D. N. & Monroe, C. Quantum interference of photon pairs from two remote trapped atomic ions. *Nature Phys.* **3**, 538–541 (2007).
24. Dehmelt, H. Radiofrequency spectroscopy of stored ions. I: Storage. *Adv. At. Mol. Phys.* **3**, 53–72 (1967).
25. Hong, C. K., Ou, Z. Y. & Mandel, L. Measurement of subpicosecond time intervals between two photons by interference. *Phys. Rev. Lett.* **59**, 2044–2046 (1987).
26. Simon, C. & Irvine, W. T. M. Robust long-distance entanglement and a loophole-free Bell test with ions and photons. *Phys. Rev. Lett.* **91**, 110405 (2003).
27. Legero, T., Wilk, T., Kuhn, A. & Rempe, G. Characterization of single photons using two-photon interference. *Adv. At. Mol. Opt. Phys.* **53**, 253–289 (2006).
28. Bollinger, J. J., Heinzen, D. J., Itano, W. M., Gilbert, S. L. & Wineland, D. J. A 303 MHz Frequency Standard based on Trapped Be^+ Ions. *IEEE Trans. Inst. Meas.* **40**, 126–128 (1991).
29. Roos, C. F. *et al.* Bell states of atoms with ultralong lifetimes and their tomographic state analysis. *Phys. Rev. Lett.* **92**, 220402 (2004).
30. Langer, C. *et al.* Long-lived qubit memory using atomic ions. *Phys. Rev. Lett.* **95**, 060502 (2005).

Acknowledgements This work is supported by the National Security Agency and the Disruptive Technology Office under Army Research Office contract, and the National Science Foundation Information Technology Research (ITR) and Physics at the Information Frontier (PIF) programmes.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.L.M. (david.moehring@mpq.mpg.de).

METHODS

Limits on excitation probability. To ensure that at most a single photon is emitted from each ion following an excitation pulse, it is important to use a laser pulse that is much shorter than the lifetime of the excited state²³. Here, we use a 2 ps laser pulse from an mode-locked, frequency-doubled Ti:sapphire laser that is much shorter than the 8.1 ns excited-state lifetime of the $^2P_{1/2}$ level. This near transform-limited pulse has a bandwidth of ~ 250 GHz which is not only much larger than the $^2P_{1/2}$ linewidth, but also much larger than the $^2S_{1/2}$ hyperfine splitting. Hence, the σ^- -polarized optical pulse that resonantly excites the $^2S_{1/2} |0, 0\rangle$ state to the $^2P_{1/2} |1, -1\rangle$ state is also resonant with the $^2S_{1/2} |1, 0\rangle \leftrightarrow ^2P_{1/2} |1, -1\rangle$ transition³¹. In this three-level lambda system, the largest population which can be transferred to the $|1, -1\rangle$ state when starting from $|0, 0\rangle$ is $P_{\text{exc}} = 50\%$, with the other 50% in an equal superposition of $|0, 0\rangle$ and $|1, 0\rangle$.

In principle, the excitation probability can be improved to unity by preparing an appropriate initial superposition of $|0, 0\rangle$ and $|1, 0\rangle$. Alternatively, different excitation schemes can be adopted¹⁷; for example, exciting to the $^2P_{1/2} |1, -1\rangle$ state from the $^2S_{1/2} |1, -1\rangle$ state via a π -polarized optical pulse.

Phase of entangled state. When considering all phases, the quantum state of the system before interference on the beam splitter is

$$\frac{1}{2} \left[\left(e^{-i\omega_1 t} |\uparrow\rangle_a e^{ik_{v_1} x_a - i\omega_{v_1} t} |v_\uparrow\rangle_a - e^{-i\omega_1 t} |\downarrow\rangle_a e^{ik_{v_1} x_a - i\omega_{v_1} t} |v_\downarrow\rangle_a \right) \otimes \left(e^{-i\omega_2 t} |\uparrow\rangle_b e^{ik_{v_2} x_b - i\omega_{v_2} t} |v_\uparrow\rangle_b - e^{-i\omega_2 t} |\downarrow\rangle_b e^{ik_{v_2} x_b - i\omega_{v_2} t} |v_\downarrow\rangle_b \right) \right] \quad (2)$$

where $\hbar(\omega_1 - \omega_\downarrow)$ and $\hbar(\omega_{v_1} - \omega_{v_\downarrow})$ are the energy differences between the two atomic and photonic qubit states, respectively, and x_i is the photon path length from the i th ion to the beam splitter. However, because $\omega_1 + \omega_{v_1} = \omega_\downarrow + \omega_{v_\downarrow}$, the equation can be rewritten as:

$$\frac{1}{2} \left[\left(e^{ik_{v_1} x_a} |\uparrow\rangle_a |v_\uparrow\rangle_a - e^{ik_{v_1} x_a} |\downarrow\rangle_a |v_\downarrow\rangle_a \right) \otimes \left(e^{ik_{v_2} x_b} |\uparrow\rangle_b |v_\uparrow\rangle_b - e^{ik_{v_2} x_b} |\downarrow\rangle_b |v_\downarrow\rangle_b \right) \right] \quad (3)$$

The two photons emerge from the beam splitter along separate paths only if they are in the antisymmetric state $|\Psi^-\rangle_{\text{photon}} = (|v_\uparrow\rangle_a |v_\downarrow\rangle_b - |v_\downarrow\rangle_a |v_\uparrow\rangle_b)/\sqrt{2}$. Upon simultaneous photon detection, the ions are projected onto

$$\frac{1}{\sqrt{2}} \left(-|\uparrow\rangle_a |\downarrow\rangle_b + e^{-i\Delta k \Delta x} |\downarrow\rangle_a |\uparrow\rangle_b \right) \quad (4)$$

where $\Delta k \equiv k_{v_1} - k_{v_2}$ and $\Delta x \equiv x_a - x_b$. We note that the entanglement is insensitive to fluctuations in the path length at the scale of the optical wavelength^{13,26}. The relative phase appearing in the entangled state of equation (4) is only

sensitive to path length fluctuations compared to the wavelength associated with the frequency difference of the photonic and atomic qubit states $2\pi/\Delta k = c/(\omega_{v_1} - \omega_{v_2}) = 2.4$ cm. Stability over this scale is easily achieved.

Fidelity and entanglement of formation. The desired resulting entangled state is $|\Psi^-\rangle_{\text{atom}} = (|\uparrow\rangle_a |\downarrow\rangle_b - |\downarrow\rangle_a |\uparrow\rangle_b)/\sqrt{2}$, so the calculated fidelity is $\mathcal{F} = (\rho_{\uparrow\downarrow, \uparrow\downarrow} + \rho_{\downarrow\uparrow, \downarrow\uparrow})/2 + |\rho_{\uparrow\downarrow, \downarrow\uparrow}|$, where $\rho_{ij,kl} = \langle ij|\rho|kl\rangle$ and $i, j, k, l \in (\uparrow, \downarrow)$. The unrotated basis measurements yield $\rho_{\downarrow\downarrow, \downarrow\downarrow}$, $\rho_{\downarrow\downarrow, \uparrow\uparrow}$, $\rho_{\uparrow\downarrow, \uparrow\downarrow}$ and $\rho_{\uparrow\uparrow, \uparrow\uparrow}$ directly. The rotated basis measurements yield $\tilde{\rho}_{\downarrow\downarrow, \downarrow\downarrow}$, $\tilde{\rho}_{\downarrow\downarrow, \uparrow\uparrow}$, $\tilde{\rho}_{\uparrow\downarrow, \uparrow\downarrow}$ and $\tilde{\rho}_{\uparrow\uparrow, \uparrow\uparrow}$, where $\tilde{\rho}_{ij,kl}$ corresponds to the density matrix elements after the applied microwave $\pi/2$ rotations with phase ϕ_a and ϕ_b on the two ions. We find

$$\tilde{\rho}_{\downarrow\downarrow, \downarrow\downarrow} + \tilde{\rho}_{\uparrow\uparrow, \uparrow\uparrow} - \tilde{\rho}_{\downarrow\uparrow, \downarrow\uparrow} - \tilde{\rho}_{\uparrow\downarrow, \uparrow\downarrow} = 2|\rho_{\downarrow\uparrow, \uparrow\downarrow}| \cos(\phi_a - \phi_b) + 2|\rho_{\downarrow\downarrow, \uparrow\uparrow}| \cos(\phi_a + \phi_b) \quad (5)$$

In the experiment, we control the relative phase $\Delta\phi = \phi_b - \phi_a$ between the microwave pulse on each ion, but have no control over the absolute phase of the applied microwaves. Therefore, the measured contrast in the rotated basis measurement comes entirely from the $\rho_{\downarrow\uparrow, \uparrow\downarrow}$ term, with a resulting fidelity $\mathcal{F} = (\rho_{\downarrow\uparrow, \uparrow\downarrow} + \rho_{\uparrow\downarrow, \uparrow\downarrow} + C)/2 = 0.63 \pm 0.03$, where C is the contrast of the oscillations in Fig. 4.

A lower bound on the entanglement of formation can be calculated by suppressing the unobserved single-qubit coherences (for example, $\rho_{\downarrow\downarrow, \uparrow\uparrow}$ or $\rho_{\uparrow\downarrow, \uparrow\uparrow}$), which cannot increase the entanglement. The resulting density matrix can then be expressed as

$$\begin{array}{c} \begin{array}{c} |\downarrow\rangle_a |\downarrow\rangle_b \\ |\downarrow\rangle_a |\uparrow\rangle_b \\ |\uparrow\rangle_a |\downarrow\rangle_b \\ |\uparrow\rangle_a |\uparrow\rangle_b \end{array} \end{array} \begin{pmatrix} 0.11 & 0 & 0 & \rho_{\downarrow\downarrow, \uparrow\uparrow} \\ 0 & 0.38 & 0.235 & 0 \\ 0 & 0.235 & 0.40 & 0 \\ \rho_{\uparrow\uparrow, \downarrow\downarrow} & 0 & 0 & 0.11 \end{pmatrix} \quad (6)$$

From this, the lower bound is numerically calculated using the procedure outlined in references³² and³³, resulting in a concurrence of $\mathcal{C} \geq 0.25 \pm 0.04$ and an entanglement of formation of $\mathcal{E} \geq 0.12 \pm 0.03$.

- Madsen, M. J. *et al.* Ultrafast coherent excitation of a trapped ion qubit for fast gates and photon frequency qubits. *Phys. Rev. Lett.* **97**, 040505 (2006).
- Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. Mixed-state entanglement and quantum error correction. *Phys. Rev. A* **54**, 3824–3851 (1996).
- Hill, S. & Wootters, W. K. Entanglement of a pair of quantum bits. *Phys. Rev. Lett.* **78**, 5022–5025 (1997).

LETTERS

Control of the electronic phase of a manganite by mode-selective vibrational excitation

Matteo Rini¹, Ra'anan Tobey², Nicky Dean², Jiro Itatani^{1,3}, Yasuhide Tomioka⁴, Yoshinori Tokura^{4,5}, Robert W. Schoenlein¹ & Andrea Cavalleri^{2,6}

Controlling a phase of matter by coherently manipulating specific vibrational modes has long been an attractive (yet elusive) goal for ultrafast science. Solids with strongly correlated electrons, in which even subtle crystallographic distortions can result in colossal changes of the electronic and magnetic properties, could be directed between competing phases by such selective vibrational excitation. In this way, the dynamics of the electronic ground state of the system become accessible, and new insight into the underlying physics might be gained. Here we report the ultrafast switching of the electronic phase of a magnetoresistive manganite via direct excitation of a phonon mode at 71 meV (17 THz). A prompt, five-order-of-magnitude drop in resistivity is observed, associated with a non-equilibrium transition from the stable insulating phase to a metastable metallic phase. In contrast with light-induced^{1–3} and current-driven⁴ phase transitions, the vibrationally driven bandgap collapse observed here is not related to hot-carrier injection and is uniquely attributed to a large-amplitude Mn–O distortion. This corresponds to a perturbation of the perovskite-structure tolerance factor, which in turn controls the electronic bandwidth via inter-site orbital overlap^{5,6}. Phase control by coherent manipulation of selected metal–oxygen phonons should find extensive application in other complex solids—notably in copper oxide superconductors, in which the role of Cu–O vibrations on the electronic properties is currently controversial.

Manganites exhibit a number of exotic phenomena, including charge-ordered and striped phases, orbital and magnetic ordering, half-metallicity, phase separation and colossal magnetoresistance^{5,6}. Most of these phenomena stem from the strong interaction between lattice, charge, orbital and spin degrees of freedom, which compete on similar energy scales to determine the ground state of the system⁷. Arguably, the most striking aspect of the physics of manganites is the occurrence of a number of metal–insulator transitions, initiated for instance via perturbations of temperature, magnetic field, pressure and irradiation with light⁵.

$\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ is a unique example among manganites, exhibiting insulating behaviour over the entire chemical composition x and over the entire temperature range⁸. This is a consequence of the small ionic radius of Ca, which results in a pronounced orthorhombic distortion (Fig. 1a) that favours charge localization⁶. Notably, the insulating phase at $x = 0.3$ adjoins a ‘hidden’ metallic state of the system, characterized by enormous changes in resistivity.

In ABO₃ perovskites, the orthorhombic distortion is quantified by the geometric ‘tolerance factor’ that depends on the average A–O (A = Pr, Ca) and B–O (B = Mn) distances:

$$\Gamma = \frac{(A - O)}{\sqrt{2}(B - O)}$$

where $\Gamma = 1$ corresponds to an ideal cube, while $\Gamma < 1$ reflects a compression of the Mn–O bond and an elongation of the A–O bond. Moreover, $\Gamma < 1$ indicates a Mn–O–Mn angle θ that is smaller than 180°, consistent with a symmetry-lowering rotation leading to orthorhombic or rhombohedral structures. The tolerance factor is related to the electronic properties of the solid via the one electron bandwidth W , because the capacity for 3d electrons to hop between neighbouring Mn-atoms depends on a super-transfer process via O(2p) states and on the degree of overlap between orbitals in neighbouring sites^{6,9,10}. The hopping matrix element reaches its maximum at $\theta = 180^\circ$ (cubic), and decreases with θ , vanishing at $\theta = 90^\circ$. Systematic studies of several $\text{A}_{0.7}\text{A}'_{0.3}\text{MnO}_3$ compounds show that the tolerance factor controls the competition between ferromagnetic metallic, paramagnetic insulating, and ferromagnetic insulating phases¹¹.

Here we show that coherent excitation of specific infrared-active modes can control the electronic phase of a manganite via direct modulation of the tolerance factor. Figure 1b shows the low-temperature optical conductivity spectrum of $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ with three dominant

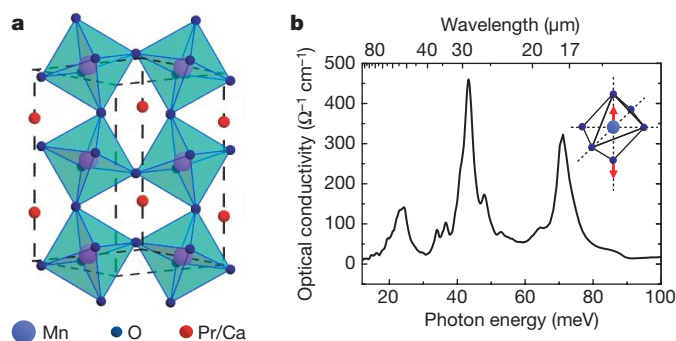


Figure 1 | $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ crystal structure and vibrational spectrum.

a, Unit cell of $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ with pronounced orthorhombic distortion resulting from the small ionic radius of the Ca atoms. The Mn–O–Mn bond is bent at an angle $\theta < 180^\circ$, which varies linearly with the tolerance factor Γ (ref. 10). The Pr/Ca doping results in an alternating network of Mn^{3+} and Mn^{4+} ions. The crystal field splits the fivefold Mn 3d levels into t_{2g} and e_g subsets. The electron hopping occurs between 3d e_g levels of neighbouring Mn^{3+} and Mn^{4+} species. The lattice distortion is related monotonically to the one electron bandwidth W , because the effective hopping interaction of 3d electrons between neighbouring Mn sites depends on super-transfer process via O(2p) states, and the p-orbital of oxygen cannot point towards two manganese atoms simultaneously if $\theta \neq 180^\circ$ (ref. 6). **b**, Low-temperature (10 K) optical conductivity spectrum of $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$. The inset shows the atomic displacements within the MnO_6 octahedra associated with the 71 meV phonon mode that modulates the Mn–O distance, and hence the tolerance factor.

¹Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ²Department of Physics, Clarendon Laboratory, University of Oxford, Oxford OX13PU, UK. ³ERATO, Japan Science and Technology Agency, Chiyoda-ku, Tokyo 102-0075, Japan. ⁴Correlated Electron Research Center, AIST, Tsukuba, Ibaraki, 305-8562 Japan. ⁵Department of Applied Physics, University of Tokyo, Tokyo 113-8656, Japan. ⁶Central Laser Facility, Rutherford Appleton Laboratory and Diamond Light Source, Chilton, Didcot, OX11 0QX, UK.

phonon modes (23, 42 and 71 meV)¹² corresponding to the three (F_{2u}) infrared active vibrational modes of a cubic perovskite. The orthorhombic distortion is responsible for the appearance of a number of weaker resonances, although only a subset of the active 25 ($7B_{1u} + 9B_{2u} + 9B_{3u}$) infrared phonon modes of a $Pbnm$ orthorhombic structure is clearly visible. The two highest-frequency vibrations are assigned to the Mn–O–Mn bending mode and the Mn–O stretching mode respectively¹³. Both vibrational modes affect the geometrical parameters determining the tolerance factor and are thus expected to have a strong coupling to the electronic properties of the system. Here we focus on the highest-frequency Mn–O stretching vibration at 71 meV and study the effect of coherent large-amplitude excitation of this mode with intense femtosecond mid-infrared pulses. The material response is investigated using both ultrafast pump–probe spectroscopy and transient conductivity measurements to characterize the insulator–metal transition^{2,3,14–17}.

$\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ single crystals with $x = 0.3$ are synthesized by a floating zone technique¹⁸ and characterized by X-ray Laue as well as neutron diffraction experiments¹⁸. The crystals are subsequently cut, polished and annealed in an oxygen environment in preparation for pump–probe studies. In the pump–probe spectroscopy studies, $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ samples at 30 K are excited by 200–300 fs, $\sim 1 \mu\text{J}$ pulses tuned in the mid-infrared spectral region around $17.5 \mu\text{m}$ (corresponding to a photon energy $h\nu = 71 \text{ meV}$). The pump pulses are focused to a fluence of about 1 mJ cm^{-2} and the transient changes in reflectivity are probed by delayed femtosecond pulses at visible to near-infrared frequencies to identify the characteristic spectral signatures and formation time of the metallic phase. Figure 2a shows the transient reflectivity $\Delta R/R$ at 800 nm after impulsive vibrational excitation at $17.5 \mu\text{m}$ and compared with above-bandgap pulsed excitation. The reflectivity responses are identical, with large long-lived changes in reflectivity developing within 1 ps of excitation. Moreover, these changes exhibit threshold and saturation dependence on the pump fluence, characteristic of a phase transformation to the metallic state, as previously established for above-bandgap excitation¹⁵.

Figure 2b shows the spectral dependence of the $\Delta R/R$ signal (at 1 ps delay) for the case of the $17.5 \mu\text{m}$ pump wavelength. The spectrum of the reflectivity changes exhibits identical features, as in previous optical studies in $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ (ref. 17), which showed that the transition to the conducting phase (induced either by applied magnetic field, or by above-bandgap transient optical excitation) is characterized by decreased reflectivity at photon energies in the 0.5–1.9 eV range and increased reflectivity at higher photon energies. Such reflectivity changes have been interpreted as a consequence of melting of the charge order and of the collapse of the 0.3 eV insulating gap, leading to the formation of a pseudo plasma edge in the metallic state¹². Our observation of the ultrafast formation of a metallic-like reflectivity spectrum after $17.5 \mu\text{m}$ pump excitation provides

evidence that the metallic state is formed promptly (within the 300 fs experimental resolution) via direct vibrational excitation, and that this state persists for hundreds of picoseconds. Although the spectral reflectivity signature is associated with the conducting phase of $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$, our recent measurements show that this is not uniquely indicative of the melting of charge order, because a similar spectral signature is observed when the metallic phase is induced by photoexcitation from the room-temperature paramagnetic insulating phase.

Figure 2c shows the dependence of the reflectivity change (measured at 1 ps delay, 800 nm probe wavelength) on the pump wavelength, in the vicinity of the phonon resonance. The observed reflectivity change clearly vanishes when the pump wavelength is tuned outside the $17.5 \mu\text{m}$ phonon absorption band. The magnitude of $\Delta R/R$ is maximum when the excitation wavelength is resonant with the Mn–O stretching mode, providing further evidence of an ultrafast vibrationally induced phase transition. The linear absorption spectrum around the $17.5 \mu\text{m}$ phonon resonance is not polarization-dependent¹² and no significant dependence on the pump polarization was observed either in the pump–probe experiments or in the conductivity measurements.

In addition to the optical measurements, changes in the sample conductivity are directly monitored by measuring the transient sample resistance after mid-infrared excitation. Gold electrodes with a $200\text{-}\mu\text{m}$ -wide gap are vacuum-evaporated onto the sample surface², and are d.c.-biased at 30 V. Measurements are performed at 30 K, where the charge-ordered, anti-ferromagnetic phase exhibits strong insulating character⁸. Laser pulses at $17.5 \mu\text{m}$ are used to excite the sample (under conditions identical to those described above), with the laser spot fully covering the space between the electrodes. The current flowing through the sample was monitored by measuring the voltage drop across a 50Ω resistor. Mid-infrared excitation results in a dramatic 1,000-fold increase in current (Fig. 3, upper panel), corresponding to a resistance drop from $2 \text{ G}\Omega$ to $1.25 \text{ M}\Omega$. The high conductivity state develops within the 4-ns resolution of the electronics and exhibits a resonance behaviour similar to that observed in the optical measurements (Fig. 2c). Figure 3 (lower panel) shows the increase in the sample conductivity derived from the measured transient resistance by assuming that the transition to the conductive state is uniform throughout the excited sample volume. Given the laser spot size at the electrodes ($200 \times 300 \mu\text{m}^2$) and the penetration depth of the mid-infrared light ($\sim 0.5 \mu\text{m}$)¹², the sample conductivity increase is estimated to exceed 10^5 , from $\sim 3 \times 10^{-8} \Omega^{-1} \text{ cm}^{-1}$ to $\sim 5 \times 10^{-3} \Omega^{-1} \text{ cm}^{-1}$. The metastable metallic phase is formed and relaxes within the experimental time resolution of 4 ns. The precise assessment of the lifetime of the metallic phase by means of transport measurements is hindered by oscillations due to ringing in the fast current detection electronics with a time-dependent sample resistance (see Fig. 3). However, the sample recovers its original resistance

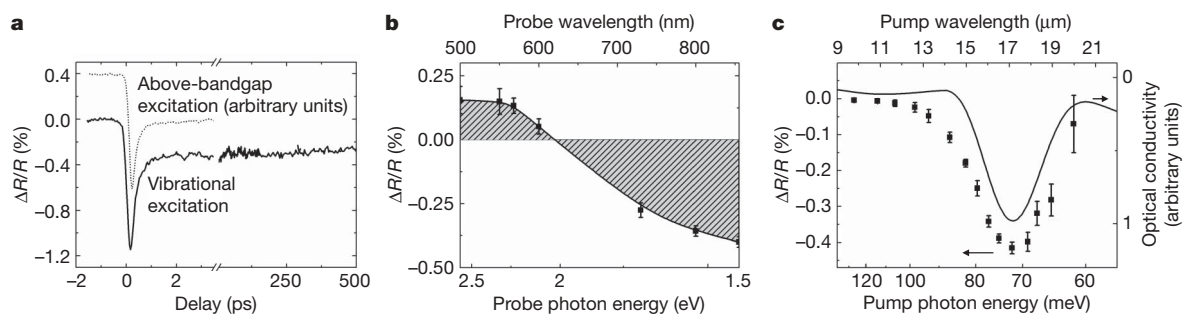


Figure 2 | Femtosecond pump–probe reflectivity studies. **a**, Relative change of reflectivity at 800 nm ($\Delta R/R$) as a function of pulse delay after vibrational excitation at $17.5 \mu\text{m}$ (solid line) and 800 nm photo-excitation (dotted line). **b**, Spectral dependence of $\Delta R/R$ measured 1 ps after vibrational excitation. The solid line is a spline fit to the data points (squares). The shift of spectral weight towards longer wavelengths, shown by the shaded area

under the curve, is a signature of the formation of the metallic phase^{12,15}. **c**, $\Delta R/R$ at 800 nm measured 1 ps after excitation (squares) and absorption spectrum around the $17.5 \mu\text{m}$ phonon resonance (solid line). For comparison, the phonon spectrum is convolved with the spectrum of the broad-bandwidth pump pulses. Error bars are s.d.

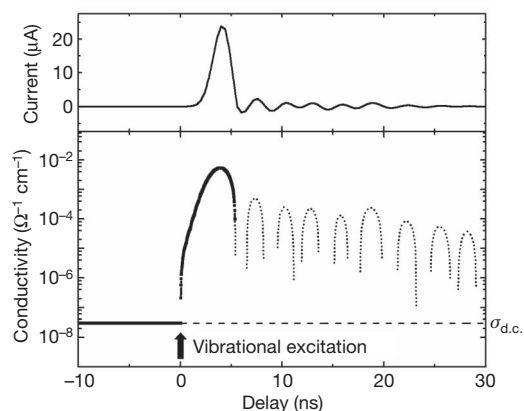


Figure 3 | Time-dependent transport measurement. Vibrational excitation of the Mn–O stretching mode results in a $\sim 10^3$ increase in the sample current (upper panel) and a corresponding $\sim 10^5$ increase in the sample conductivity (lower panel). The metastable metallic phase is formed and relaxes within the experimental time resolution of 4 ns. The current oscillations following the main pulse are due to electronic ringing and cannot be converted accurately into sample conductivity, so the derived conductivity oscillations are shown as a dotted line. The dashed line shows the d.c. conductivity of the insulating phase of $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ at 30 K.

within several tens of nanoseconds. In these measurements, contributions from interband carrier excitations are negligible because five photons at $17.5\ \mu\text{m}$ (71 meV) are required to span the 0.3-eV insulating bandgap of $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ (ref. 12). The moderate temperature jump due to laser excitation (estimated at $< 2\ \text{K}$ for excitation at the peak of the phonon resonance) can also be ruled out as the origin of the resistivity drop. As emphasized in the introduction, in $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ an insulator-to-metal phase transition cannot be induced by temperature⁸.

Our results clearly show that resonant excitation of the Mn–O phonon vibration in $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ drives the system on a femtosecond timescale into a metastable, nanosecond-lived, high-conductivity phase. We note that this occurs in the electronic ground state of the solid and no electronic excitation is involved. These results strongly suggest that coherent modulation of the ‘tolerance factor’ gives rise to dramatic changes in the electron hopping probability, providing important new insight into the physics underlying the behaviour of this strongly correlated material.

Given the present limitations in generating mid-infrared pulses beyond $25\ \mu\text{m}$, it is not yet possible to assess the specificity of the Mn–O stretching vibration by comparing it with excitation of lower-frequency phonon modes, including the Mn–O–Mn bending mode, which is also likely to modulate the tolerance factor significantly. Moreover, an important role in inducing the phase transition and subsequently stabilizing the metallic phase may be played by ultrafast vibrational energy redistribution via anharmonic coupling to the Mn–O–Mn bend, or to other modes (for example, non-infrared-active Jahn–Teller modes¹⁹) that may also influence the electron localization and delocalization.

In future, experimental techniques such as magneto-optic Kerr effect spectroscopy or time-resolved photo-emission electron microscopy²⁰ will be important for probing the spin and magnetization dynamics after mid-infrared excitation. In a magnetoresistive manganite metallicity is associated with ferromagnetism through the double-exchange mechanism²¹, so the formation of a metallic state implies the possibility of generating ferromagnetic domains on ultrafast timescales by excitation of specific vibrational degrees of freedom—a remarkable consequence of the strong coupling between magnetic, electronic and lattice degrees of freedom.

In summary, we have demonstrated the excitation of a specific phonon mode as a tool to drive the solid in the electronic ground

state towards a competing phase of the system. The ultrafast vibrational control of correlated-electron phases is probably applicable in other interesting cases, opening a new window on their controversial physics and enabling time-resolved measurements under the unique conditions created by the initial localization of energy in specific vibrational modes. This approach may extend well beyond the case of colossal magnetoresistance manganites, providing new insight into the behaviour of complex matter, including the controversial nature of high-transition-temperature superconductivity and the role played by lattice vibrations in determining its electronic properties.

Received 5 May; accepted 23 July 2007.

- Kiriukhin, V. *et al.* An X-ray induced insulator–metal transition in a magnetoresistive manganite. *Nature* **386**, 813–815 (1997).
- Miyano, K., Tanaka, T., Tomioka, Y. & Tokura, Y. Photoinduced insulator-to-metal transition in a perovskite manganite. *Phys. Rev. Lett.* **78**, 4257–4260 (1997).
- Fiebig, M., Miyano, K., Tomioka, Y. & Tokura, Y. Visualization of the local insulator–metal transition in $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$. *Science* **280**, 1925–1928 (1998).
- Asamitsu, A., Tomioka, Y., Kuwahara, H. & Tokura, Y. Current switching of resistive states in magnetoresistive manganites. *Nature* **388**, 50–52 (1997).
- Tokura, Y. *Colossal Magnetoresistive Oxides* Ch. 1 (Gordon & Breach Science Publishers, Amsterdam, 2000).
- Dagotto, E. *Nanoscale Phase Separation and Colossal Magnetoresistance* Ch. 2–3 (Springer, Berlin/Heidelberg/New York, 2002).
- Salamon, M. B. & Jaime, M. The physics of manganites: structure and transport. *Rev. Mod. Phys.* **73**, 583–628 (2001).
- Tomioka, Y., Asamitsu, A., Kuwahara, H. & Moritomo, Y. Magnetic-field-induced metal–insulator phenomena in $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ with controlled charge-ordering instability. *Phys. Rev. B* **53**, 1689–1692 (1996).
- Anderson, P. W. & Hasegawa, H. Considerations on double exchange. *Phys. Rev.* **100**, 675–681 (1955).
- Imada, M., Fujimori, A. & Tokura, Y. Metal–insulator transitions. *Rev. Mod. Phys.* **70**, 1039–1263 (1998).
- Hwang, H. Y., Cheong, S.-W., Radaelli, P. G., Marezio, M. & Batlogg, B. Lattice effects on the magnetoresistance in doped LaMnO_3 . *Phys. Rev. Lett.* **75**, 914–917 (1995).
- Okimoto, Y., Tomioka, Y., Onose, Y., Otsuka, Y. & Tokura, Y. Optical study of $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ ($x=0.4$) in a magnetic field: Variation of electronic structure with charge ordering and disordering phase transitions. *Phys. Rev. B* **59**, 7401–7408 (1999).
- Boris, A. V. *et al.* Infrared optical properties of $\text{La}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$ epitaxial films. *J. Appl. Phys.* **81**, 5756–5758 (1997).
- Ogawa, K., Wei, W., Miyano, K., Tomioka, Y. & Tokura, Y. Stability of a photoinduced insulator–metal transition in $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$. *Phys. Rev. B* **57**, R15033–R15036 (1998).
- Fiebig, M., Miyano, K., Tomioka, Y. & Tokura, Y. Reflection spectroscopy on the photoinduced local metallic phase of $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$. *Appl. Phys. Lett.* **74**, 2310–2312 (1999).
- Fiebig, M., Miyano, K., Satoh, T., Tomioka, Y. & Tokura, Y. Action spectra of the two-stage photoinduced insulator–metal transition in $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$. *Phys. Rev. B* **60**, 7944–7949 (1999).
- Fiebig, M., Miyano, K., Tomioka, Y. & Tokura, Y. Sub-picosecond photo-induced melting of a charge-ordered state in a perovskite manganite. *Appl. Phys. B* **71**, 211–215 (2000).
- Tomioka, Y., Asamitsu, A., Moritomo, Y. & Tokura, Y. Anomalous magnetotransport properties of $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$. *J. Phys. Soc. Jpn* **64**, 3626–3630 (1995).
- Millis, A. J., Shraiman, B. I. & Mueller, R. Dynamic Jahn–Teller effect and colossal magnetoresistance in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. Lett.* **77**, 175–178 (1996).
- Choe, S.-B. *et al.* Vortex core-driven magnetization dynamics. *Science* **304**, 420–422 (2004).
- Zener, C. Interaction between the *d*-shells in the transition metals. II. Ferromagnetic compounds of manganese with perovskite structure. *Phys. Rev.* **82**, 403–405 (1950).

Acknowledgements We thank Y. Okimoto for providing the optical conductivity spectra and S. Wall for help in the figures preparation. This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, of the US Department of Energy. Work at the University of Oxford, UK, was supported by the European Science Foundation through a European Young Investigator Award, and by the Oxford University Press through a John Fell Award.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.R. (mrini@ibl.gov) and to A.C. (a.cavalleri1@physics.ox.ac.uk).

The potential for giant tsunamigenic earthquakes in the northern Bay of Bengal

Phil R. Cummins¹

The great Sumatra–Andaman earthquake and Indian Ocean tsunami of 2004 came as a surprise to most of the earth science community. Although it is now widely recognized that the risk of another giant earthquake is high off central Sumatra¹, just east of the 2004 earthquake, there seems to be relatively little concern about the subduction zone to the north, in the northern Bay of Bengal along the coast of Myanmar. Here I show that similar indicators suggest a high potential for giant earthquakes along the coast of Myanmar. These indicators include the tectonic environment, which is similar to other subduction zones that experience giant megathrust earthquakes, stress and crustal strain observations, which indicate that the seismogenic zone is locked, and historical earthquake activity, which indicates that giant tsunamigenic earthquakes have occurred there in the past. These are all consistent with active subduction in the Myanmar subduction zone and I suggest that the seismogenic zone extends beneath the Bengal Fan. I conclude therefore that giant earthquakes probably occur off the coast of Myanmar, and that a large and vulnerable population is thereby exposed to a significant earthquake and tsunami hazard.

The tectonic setting of west Myanmar has been marked by considerable uncertainty and some controversy^{2,3}. Research in the 1980s^{4,5} inferred that the surface expression of the boundary between the Indian plate and southeast Asia follows the northward extension of the Andaman trough, coming ashore on the coast of Myanmar just south of Ramree island at roughly 20° N latitude (Fig. 1a). This coincides with the extrapolation of the Wadati Benioff zone beneath western Myanmar to the western flank of the Indo-Burma ranges, and is the plate boundary indicated in most widely used plate models (Fig. 1b). Most authors^{3–5} favoured the interpretation that eastward subduction of the Indian plate has ceased and the slab beneath Burma is being dragged northward along with the Indian plate. This conclusion seems to be supported by the fact that the focal mechanisms of shallow earthquakes near the presumed plate boundary have pressure axes oriented north–south, implying no east–west compression and hence no active subduction. Even if subduction were active, this position of the plate boundary suggests that convergence is highly oblique (<5°), implying a very long recurrence interval for thrust earthquakes, and that the offshore segment of the plate boundary north of the 2004 Sumatra–Andaman earthquake's rupture area would be only 300 km in length; hence, the likelihood of a giant tsunamigenic earthquake in the northern Bay of Bengal would be low⁶.

The up-dip limit of the seismogenic zone should not, however, be presumed to lie where the slab begins to descend beneath the upper plate. Rather, the seismogenic zone extends as far out along the decollement as necessary to encompass all of the significant strain accumulation that might be released in a megathrust earthquake. Although this distinction may not be important for most other

subduction zones, which have at most a 2-km-thick sediment layer on the upper surface of the subducting plate⁷, the northern Bay of Bengal is unique because it hosts the world's largest submarine fan system, the Bengal Fan, consisting of sediments that have been shed off Tibet and the Himalayas since the Early Miocene. The thickness of the Bengal Fan sediments reaches up to 20 km (ref. 8). Because even a 1-km-thick sediment cover can insulate the underlying plate enough to cause significant up-dip extension of the thermal regime required for seismogenesis⁹, it seems reasonable to expect this effect to be particularly pronounced for the extreme thickness of the Bengal Fan sediments. Such considerations support an alternative hypothesis for subduction along the coast of Myanmar: that subduction is active and “rare great earthquakes may occur along a basal decollement”⁴. This hypothesis is supported by studies in the 1990s that place the deformation front well offshore of the Myanmar coast, not coming ashore until the Mehgna estuary at 23° N (refs 8, 10–12). Marine seismic surveys¹² demonstrate unequivocally that the deformation front is offshore of Ramree island, and borehole breakout data show that the stress field in the sedimentary pile of the Bengal basin is oriented east–west¹³, suggesting that the Chittagong–Tripura fold belt is an active feature forming in response to ongoing subduction of the Indian plate beneath southeast Asia¹².

This extension of the seismogenic zone beneath the Bengal Fan, 100–200 km westward of the previously inferred plate boundary, has important implications for seismic and tsunami hazard. First, the convergence is much less oblique than it would be along the northward extension of the Andaman trench (Fig. 1b), suggesting that recurrence times of large thrust earthquakes might be only a few hundred years¹⁴. Second, it means that the shallow earthquakes in the Indo-Burma ranges with focal mechanisms exhibiting north–south pressure axes are actually occurring in the forearc, 100 km landward from the deformation front. Both of these factors would make the Arakan subduction zone similar to the Cascadia and Nankai subduction zones, which have oblique, partitioned subduction, arc-parallel orientation of the maximum horizontal stress in the forearc¹⁵, and exhibit very little megathrust earthquake activity (apart from the relatively infrequent occurrence of giant megathrust earthquakes). Thick trench sediments, which in the Bay of Bengal far exceed those of any other trench, have been shown to have a positive correlation with subduction zone earthquake occurrence¹⁶. Finally, extension of the seismogenic zone to offshore Myanmar means that the entire 900-km-long seismogenic zone, from the northern tip of the Andaman Trench to the northern tip of the Bay of Bengal, is submarine, and therefore capable of hosting a giant tsunamigenic earthquake.

Along with the interpretation that there is no active subduction between the Indian plate and southeast Asia, it has been suggested³ that all of the relative motion between the Indian and Eurasian plates is accommodated along the Sagaing fault in central Myanmar.

¹Geoscience Australia, GPO Box 378, Canberra ACT 2601, Australia.

Global positioning system (GPS) surveys^{14,17}, however, have shown that only 60% of the oblique component of relative plate motion is accommodated on the Sagaing fault. The remaining deformation could be explained either by distributed deformation west of the Sagaing fault, or by locking of the Arakan subduction zone. In the latter case, the Arakan subduction zone “would be expected to produce a magnitude 8.5 earthquake every century or a magnitude 9 every 500 years”¹⁴.

Many of the above studies have referred to historical earthquakes near the Myanmar coast, but none has considered in detail the historical reports of the 2 April 1762, Arakan earthquake. Descriptions of the effects of this earthquake and associated ground movements along the Arakan coast were recorded during a survey in 1841 by the British ship *Childers*¹⁸, commanded by Captain E. Halsted. Halsted recorded evidence of 3–7 m of uplift along the coasts of Ramree, Cheduba and Foul islands, which lie offshore of and parallel to the Arakan coast of Myanmar. Halsted remarked that the traces of this sudden change in land elevation were “as clear as could be wished”, and that on Cheduba “the natives are all perfectly aware of the bank having formerly been the limit of their island”. An interview with a local who experienced the 1762 earthquake, who “had been accustomed to fish over the now-upraised land”, confirmed that the uplift occurred during the earthquake. Halsted also noted the existence of at least one additional upraised terrace. Of particular concern for hazard assessment is Halsted’s statement that “these elevations are considered periodical by the inhabitants, occurring every hundred years, and the next event is expected within the course of the next few years, and would excite but little surprise”.

Did the 1762 earthquake cause a tsunami? Halsted’s interviews with the inhabitants of Cheduba confirm that there was a local tsunami: “the sea washed to and fro several times with great fury, and then retired from the ground”. Remarkably, no lives were reported

lost, and it is interesting to speculate that local knowledge of the tsunami threat caused inhabitants of Cheduba to flee to high ground when the earthquake occurred. Whether a tsunami affected other parts of the Bay of Bengal is less clear. There are no known reports from the coast of Bengal (now Bangladesh and the Indian province of West Bengal). At Dhaka, however, the river rose suddenly and “hundreds of large country boats were driven ashore or lost, and great numbers of lives lost with them”¹⁹. It is not clear, however, if this was due to a wave from the sea or a change in land level that may have caused the river to back up.

The geological evidence of uplift along the Arakan coast provides clear evidence of the occurrence of large thrust earthquakes there, and Halsted’s account indicates that rupture from the 1762 earthquake extended as far south as Foul island. It is less clear how far north rupture from this earthquake extended. On the one hand, there are no reports of geologic evidence for uplift north of Ramree island. On the other hand, contemporary accounts¹⁹ from Chittagong describe widespread liquefaction and subsidence. Although in many cases submergence due to the expulsion of groundwater is difficult to distinguish from inundation by the sea, there are also accounts of islands sinking to or below sea level, and a report that 60 square miles of land were “permanently submerged”²⁰, so it seems clear that at least some parts of the coast near Chittagong subsided.

This subsidence is a crucial observation, because it implies that fault rupture extended as far north as Chittagong, and that the up-dip limit of rupture was well offshore, because subsidence would occur only above the down-dip limit of fault rupture. The extension of the rupture at least as far north as Chittagong is supported by contemporary accounts describing intense ground shaking, extensive damage to buildings, and formation of large fissures and landslides¹⁹, which are all observations consistent with a Modified Mercalli Intensity (MMI) of 9 or greater. Studies of fault rupture associated with historical subduction zone earthquakes in Japan²¹, as well as

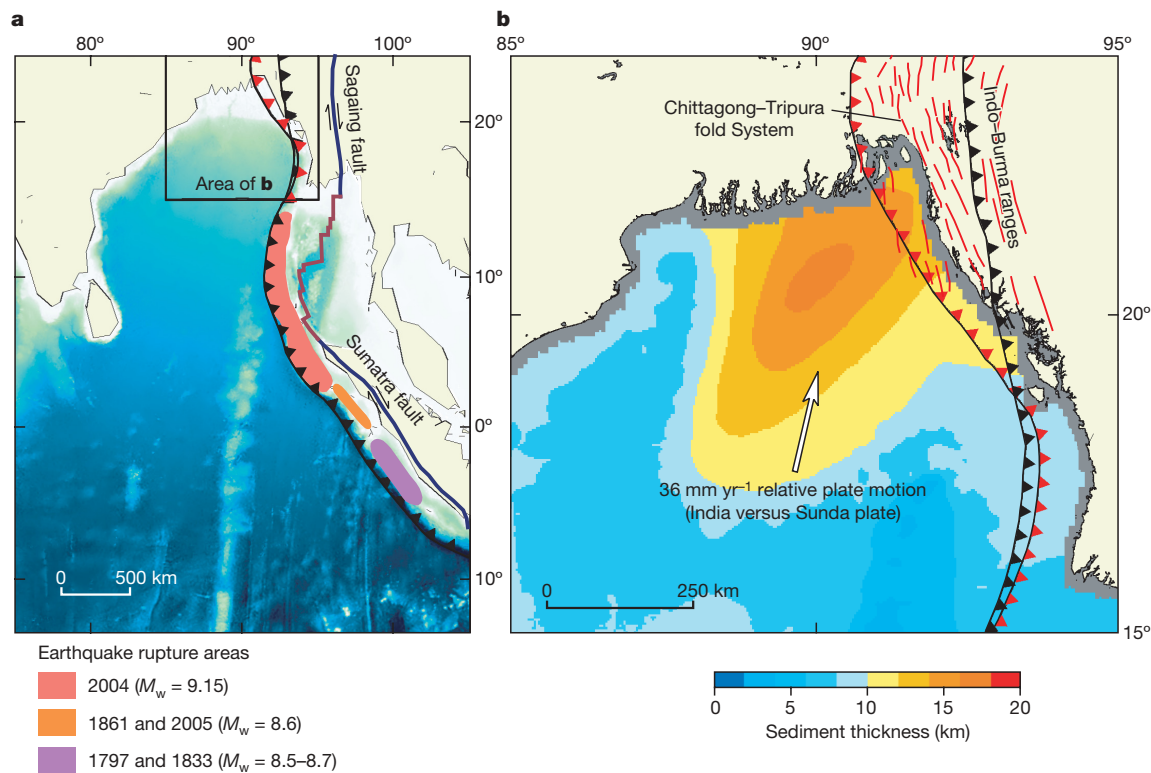


Figure 1 | Tectonic setting of the Bay of Bengal. **a**, Tectonic setting of the eastern Indian Ocean, showing Sumatra earthquake rupture areas, the subduction zone plate boundary (line with black triangles) from the University of Texas Institute of Geophysics plates model, and the deformation front¹¹ (line with red triangles). **b**, A more detailed view of the

northern Bay of Bengal, showing the rough positions of the Chittagong-Tripura fold belt and Indo-Burma ranges with respect to the plate boundary (line with black triangles) and deformation front (line with red triangles). Sediment thickness is also indicated, along with the relative motion of the Indian plate versus the Sunda plate¹⁴.

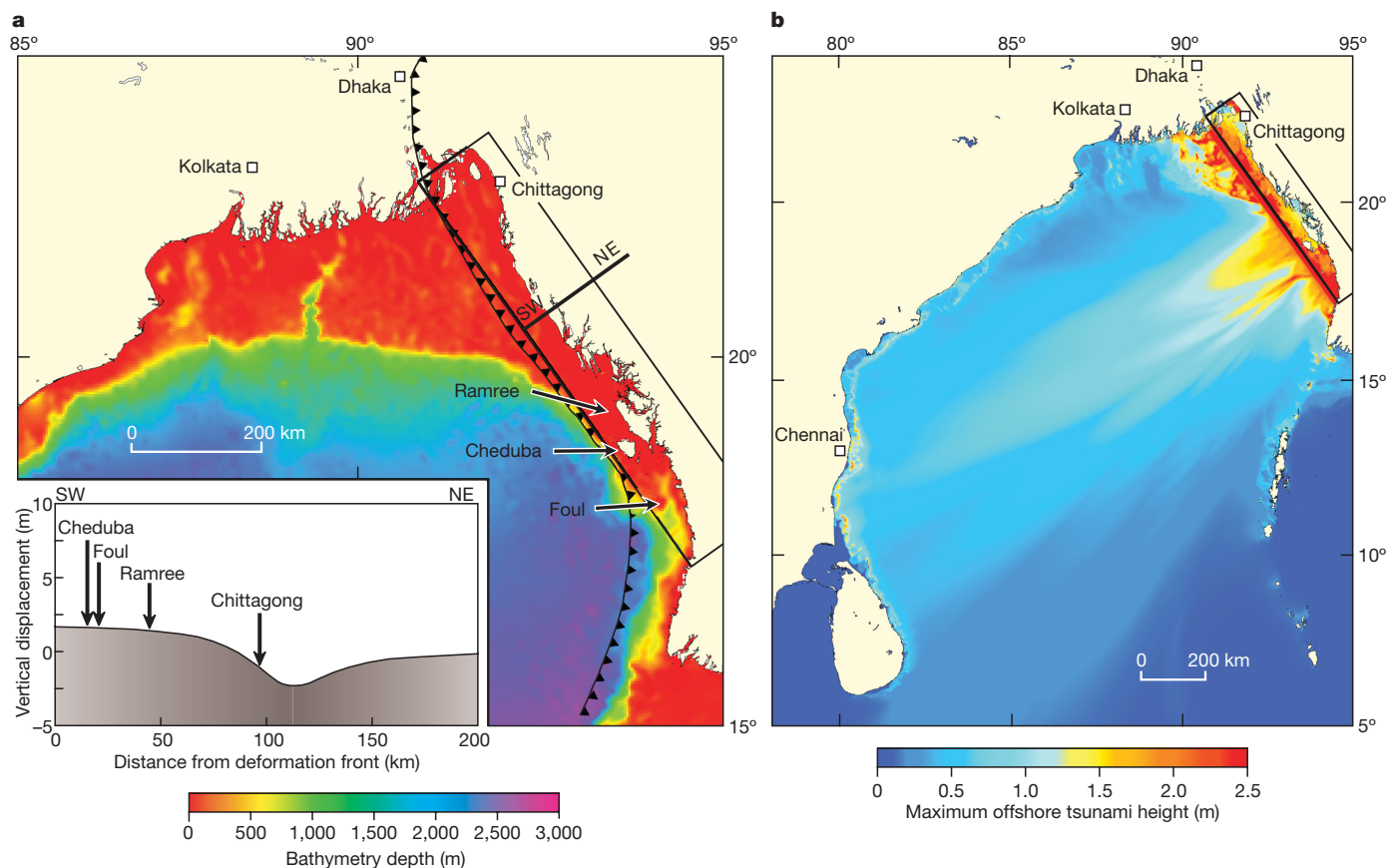


Figure 2 | Models for the 1762 Arakan earthquake and tsunamis. **a**, A fault model for the 1762 Arakan earthquake, with length 700 km, width 125 km (indicated by the rectangle along the eastern coast of the northern Bay of Bengal) and slip 10 m. Parameters were chosen to reproduce roughly the observed subsidence and uplift associated with the 1762 Arakan earthquake, with the fault's upper edge coincident with the deformation front. The inset

is a southwest–northeast (SW–NE) profile of the vertical displacement across the fault, with arrows indicating the down-dip positions of the observed uplift and subsidence. **b**, Maximum offshore heights of a tsunami, calculated using the source in **a**. This calculation does not take into account run-up, which can be many times greater than the offshore heights.

studies of macroseismic data in Mexico²², suggest that such high MMIs are only observed within 50 km of fault rupture. Finally, if rupture had extended only as far north as Ramree island, the rise of the river at Dhaka could be attributable only to a tsunami that had propagated along the strike of the subduction zone. Little tsunami energy would be propagated in this direction, so it seems unlikely that a tsunami large enough to travel 100 km upriver to Dhaka would have been generated by an earthquake whose rupture stopped at Ramree island.

A tsunami was simulated for a moment magnitude $M_w = 8.8$ thrust earthquake along the Arakan coast, positioned so that its upper edge lies along the deformation front¹¹ and 10 m reverse slip results in the observed pattern of uplift and subsidence (Fig. 2a). As discussed above, this pattern requires the seismogenic zone to extend offshore, beneath the Bengal Fan, consistent with geologic interpretations and stress observations. Such an earthquake would generate a large tsunami that could have a pronounced impact on the Chittagong coast and the Ganges–Bhramaputra delta at the northern tip of the Bay of Bengal (Fig. 2b). The latter region is home to over 60 million people living within 10 m of sea level. Chittagong, with a population of over 6 million, lies just above the presumed fault, and is the source of some of the most dramatic accounts of ground deformation and damage from the 1762 earthquake. Dhaka and Kolkata, although they are some distance from the presumed rupture area, are urban agglomerations of over 10 million each, with a large proportion of their populations living in structures that are unlikely to be seismically resilient. Even if only a small fraction of this population is vulnerable to a giant tsunamigenic earthquake along the

Arakan subduction zone, it seems likely that the number of lives at risk may be over a million.

The $M_w = 8.8$ earthquake proposed here as representative of the 1762 event is a worst-case scenario, and the details of the rupture model may be open to question. Also, it may be over 200 years before a similar event occurs, according to GPS observations of 23 mm yr^{-1} accumulation of backslip along the Arakan trench¹⁴. The next tsunamigenic earthquake in the Arakan subduction zone may not be this large, but it may also occur much sooner. It is also not certain that the 1762 earthquake produced more than a local tsunami, although it should be borne in mind that the rapid rate of sedimentation in the Bay of Bengal could also lead to tsunamis caused by submarine landslides. In any case, the evidence of active convergence accompanied by thrust earthquake activity along a coastal region with an extremely high population density suggests that the risk of a major tsunami in the northern Bay of Bengal should be taken seriously.

Received 27 February; accepted 10 July 2007.

1. Natawidjaja, D. H. *et al.* Source parameters of the great Sumatra megathrust earthquakes of 1797 and 1833 inferred from coral microatolls. *J. Geophys. Res.* **111**, doi:10.1029/2005JB004025 (2006).
2. Satyabala, S. P. Subduction in the Indo-Burma region: Is it still active? *Geophys. Res. Lett.* **25**, 3189–3192 (1998).
3. Guzmán-Speziale, M. & Ni, J. F. Comment on “Subduction in the Indo-Burmese region: Is it still active?” by S. P. Satyabala. *Geophys. Res. Lett.* **27**, 1065–1066 (2000).
4. Le Dain, A. Y., Tapponier, P. & Molnar, P. Active faulting and tectonics of Burma and the surrounding regions. *J. Geophys. Res.* **89**, 453–472 (1984).
5. Ni, J. F. *et al.* Accretionary tectonics of Burma and the three-dimensional geometry of the Burma subduction zone. *Geology* **17**, 68–71 (1989).

6. Kamesh Raju, K. A., Murty, G. P. S., Amarnath, D. & Mohan Kumar, M. L. The west Andaman fault and its influence on the aftershock pattern of the recent megathrust earthquakes in the Andaman-Sumatra region. *Geophys. Res. Lett.* **34**, doi:10.1029/2006GL028730 (2007).
7. Scholl, D. Sediment influx to subduction zone and great magnitude earthquakes; an observational appraisal. In *Tsunami Sources Workshop Proceedings* (US Geological Survey, Menlo Park, California, 21–22 April 2006).
8. Alam, M., Alam, M. M., Curray, J. R., Rahman Chowdhury, M. L. & Royhan Gani, M. An overview of the sedimentary geology of the Bengal Basin in relation to the regional tectonic framework and basin-fill history. *Sedim. Geol.* **155**, 179–208 (2003).
9. Wang, K., Hyndman, R. D. & Yamamoto, M. Thermal regime of the southwest Japan subduction zone: effects of age history of the subducting plate. *Tectonophysics* **248**, 53–69 (1995).
10. Kahn, A. A. Tectonics of the Bengal Basin. *J. Himal. Geol.* **2**, 91–101 (1991).
11. Acharyya, S. Break-up of the greater Indo-Australian continent and accretion of blocks framing south and east Asia. *J. Geodyn.* **26**, 149–170 (1998).
12. Nielson, C., Chamot-Rooke, N., Rangin, C. & the ANDAMAN Cruise Team. the ANDAMAN Cruise Team. From partial to full strain partitioning along the Indo-Burmese hyper-oblique subduction. *Mar. Geol.* **209**, 303–327 (2004).
13. Gowd, T. N., Sri Rama Rao, S. V. & Gaur, V. K. Tectonic stress field in the Indian subcontinent. *J. Geophys. Res.* **97**, 11879–11888 (1992).
14. Socquet, A. *et al.* India and Sunda plates motion and deformation along their boundary in Myanmar determined by GPS. *J. Geophys. Res.* **111**, doi:10.1029/2005JB003877 (2006).
15. Wang, K. & He, J. Mechanics of low stress forearcs: Nankai and Cascadia. *J. Geophys. Res.* **104**, 15191–15206 (1999).
16. Ruff, L. Do trench sediments affect great earthquake occurrence in subduction zones? *Pure Appl. Geophys.* **129**, 263–282 (1989).
17. Vigny, C. *et al.* Present-day crustal deformation around Sagaing fault, Myanmar. *J. Geophys. Res.* **108**, doi:10.1029/2002JB001999 (2003).
18. Halsted, E. P. Report on the Island of Cheduba. *J. Asiatic Soc. Bengal* **114**, 319–446 (1843).
19. Hirst, W. An Account of an Earthquake at Chattigaon: translated from the Persian by Mr. Edward Gulston, in the Service of the Honourable East India Company, and Communicated by Him to the Reverend Mr. Hirst. *Phil. Trans. R. Soc. Lond.* **8**, 251 (1763).
20. Oldham, T. A catalogue of Indian earthquakes from the earliest time to the end of A.D. 1869. *Mem. Geol. Surv. India* **19**, 163–215 (1883).
21. Ando, M. Source mechanisms and tectonic significance of historical earthquakes along the Nankai Trough, Japan. *Tectonophysics* **27**, 119–140 (1975).
22. Chavez, M. & Castro, R. Attenuation of Modified Mercalli Intensity with distance in Mexico. *Bull. Seismol. Soc. Am.* **78**, 1875–1884 (1988).

Acknowledgements Among the many colleagues at Geoscience Australia who offered support and advice, I especially thank C. Collins for discussions about the 1762 Arakan earthquake and Bengal during the colonial era, and M. Hollow for help locating many of the historical references. I am also grateful to K. Sieh and C. Vigny for their thorough and constructive reviews.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.R.C. (phil.cummins@ga.gov.au).

Raptorial jaws in the throat help moray eels swallow large prey

Rita S. Mehta¹ & Peter C. Wainwright¹

Most bony fishes rely on suction mechanisms to capture and transport prey¹. Once captured, prey are carried by water movement inside the oral cavity to a second set of jaws in the throat, the pharyngeal jaws, which manipulate the prey and assist in swallowing^{1,2}. Moray eels display much less effective suction-feeding abilities³. Given this reduction in a feeding mechanism that is widespread and highly conserved in aquatic vertebrates, it is not known how moray eels swallow large fish and cephalopods^{4–7}. Here we show that the moray eel (*Muraena retifera*) overcomes reduced suction capacity by launching raptorial pharyngeal jaws out of its throat and into its oral cavity, where the jaws grasp the struggling prey animal and transport it back to the throat and into the oesophagus. This is the first described case of a vertebrate using a second set of jaws to both restrain and transport prey, and is the only alternative to the hydraulic prey transport reported in teleost fishes. The extreme mobility of the moray pharyngeal jaws is made possible by elongation of the muscles that control the jaws⁸, coupled with reduction of adjacent gill-arch structures⁹. The discovery that pharyngeal jaws can reach up from behind the skull to grasp prey in the oral jaws reveals a major innovation that may have contributed to the success of moray eels as apex predators hunting within the complex matrix of coral reefs^{10,11}. This alternative prey transport mode is mechanically similar to the ratcheting mechanisms used in snakes^{12,13}—a group of terrestrial vertebrates that share striking morphological, behavioural¹⁴ and ecological convergence with moray eels.

Anguilliform fishes, collectively known as eels, are characterized by an elongate body, a small cross-sectional area and an absence of pelvic fins¹⁵. These characteristics enable eels to live in confined spaces, promoting reclusive and enigmatic lifestyles^{10,16}. Within the Anguilliformes, moray eels (Muraenidae) are a monophyletic group of predators that occupy coral crevices of shallow reefs^{10,11,16}. Despite high species richness (roughly 200 species constitute approximately one-quarter of eel diversity worldwide¹¹) and their role as top predators of many coral reef ecosystems^{7,17,18}, the mechanisms of feeding in morays are poorly understood, particularly with respect to the large prey they are known to eat^{4–7}. Given the striking morphological and ecological features of morays,

understanding the basis of their feeding performance may provide insight into their successful radiation on coral reefs.

In a recent study, we described the feeding kinematics of two moray species with diverse dietary habits, calling attention to the marked reduction of important suction-producing cranial elements, particularly the hyoid skeleton and sternohyoideus muscle³. From kinematic analysis of feeding behaviour, we concluded that morays do not use suction to capture prey but, rather, apprehend prey by biting. Biting and suction are not mutually exclusive mechanisms¹⁹. Among ray-finned fishes, the use of a suction-induced flow of water is the only known mechanism for transporting prey from the oral jaws to the pharyngeal jaws and oesophagus^{1,20}.

Morays have a well-developed pharyngeal jaw apparatus that is positioned posterior to the skull—a more caudal position compared to the pharyngeal jaws of other teleosts^{9,21} (Fig. 1a). In light of their reduced capacity to suction water with their jaws, we explored the possibility that morays have evolved an alternative to hydraulic-based prey transport to move large prey from the oral jaws to the pharyngeal jaws, a distance that is secondarily elongate in morays⁹.

We used high-speed video to study intra-oral transport behaviour in reticulated morays, *M. retifera*. In each feeding sequence, once the prey was captured in the oral jaws, morays protracted their pharyngeal jaws forward to ensnare the prey, which was then pulled into the moray's throat (see Supplementary Movies 1 and 2). In 40 intra-oral transport sequences, morays protracted their pharyngeal jaws into the oral cavity 88% of the time. Pharyngeal jaw protraction extended as far as the anterior margin of the orbit, a total of 3.5 cm in an eel with a total head length of 3.3 cm (Fig. 1b). In the remaining four trials, we observed anterior–posterior movement of the gill-arch area, but the pharyngeal jaws did not extend past the corner of the mouth. When the pharyngeal jaws engage the prey item, morays increase their oral jaw gape, releasing the prey and extend their head forward while the pharyngeal jaws retract, dragging the prey into the oesophagus (Fig. 2). This mechanism of prey transport is behaviourally and functionally convergent to prey transport in snakes. Snakes transport prey by alternating ratcheting movements of the left and right sides of their upper jaws to advance their head over prey^{12,13}.

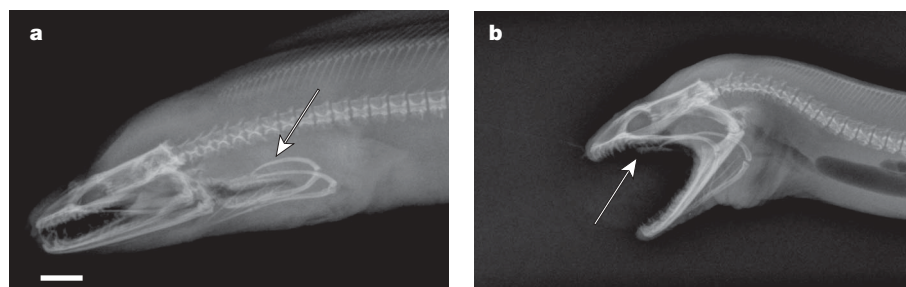


Figure 1 | Radiographs in left lateral view depicting the extreme positions of the pharyngeal jaws in *M. retifera* during prey transport. a, Posterior placement of the pharyngeal jaws in relation to the skull. The arrow points to the pharyngeal jaws. **b**, Pharyngeal jaws in their protracted position. The arrow points to the upper pharyngobranchial. Scale bar for **a** and **b**, 1 cm.

¹Section of Evolution and Ecology, University of California, One Shields Avenue, Davis, California 95616, USA.

Oral jaw ratcheting is coupled with flexion and extension of their cervical and thoracic vertebrae¹³. Ratcheting enables snakes to maintain a bite on their prey with one side of their upper jaws while the opposing side moves over the prey. Morays transport prey using alternating movements of the oral and pharyngeal jaws followed by flexion and extension of the anterior region of the vertebral column. The dual-jaw system of morays functions to alternate oral biting with pharyngeal biting, allowing morays to maintain a grip on their prey at all times, as snakes do.

Pharyngeal jaws are modified gill arches that are bilaterally paired^{2,22}. In many teleost fishes, the upper pharyngeal jaws are broad plates. Both upper and lower pharyngeal jaws bear teeth. The lower pharyngeal jaws press against the upper pharyngeal jaws while the latter are moved posteriorly in a shearing motion to manipulate material that is transported into the throat, a function that is relatively conserved across bony fish^{2,23}. The upper and lower pharyngeal jaws of *M. retifera* are not broad structures, but rather elongate, thin, grasping arms. The upper arms are formed by the fourth epibranchials and the lower are formed by the fourth ceratobranchials. The upper and lower jaws bear sharp recurved teeth, giving the impression of talons (Fig. 3a, b). A hinge, which attaches the upper pharyngobranchial to the epibranchial, enables the toothed pharyngobranchial to rotate dorsally. The upper jaws have slight independent anterior–posterior-directed movement and greater lateral movement. The left and right sides of the lower pharyngobranchials are joined anteriorly by a region of connective tissue fibres and are restricted to anterior–posterior-directed movement. The design of the moray pharyngeal jaw represents specialization for extreme transport movements to carry prey from the oral jaws into the oesophagus.

In most groups of bony fishes, the pharyngeal jaws are suspended from the neurocranium, just posterior to the orbit and rostral to the oesophagus. The jaws are further bracketed anteriorly by the first to third gill-arch elements and posteriorly and ventrally by the pectoral girdle. These larger skeletal elements surrounding the pharyngeal apparatus stabilize and limit pharyngeal motion, which is powered by short protractor and retractor muscles². The extreme range of pharyngeal jaw motion in *M. retifera* is made possible by elongation of the pharyngeal muscles and reduction in the anterior branchial elements that constrain the pharyngeal movements in other bony fish.

A videofluoroscopy sequence coupled with anatomic dissections revealed the pharyngeal jaws in their full range of motion and enabled

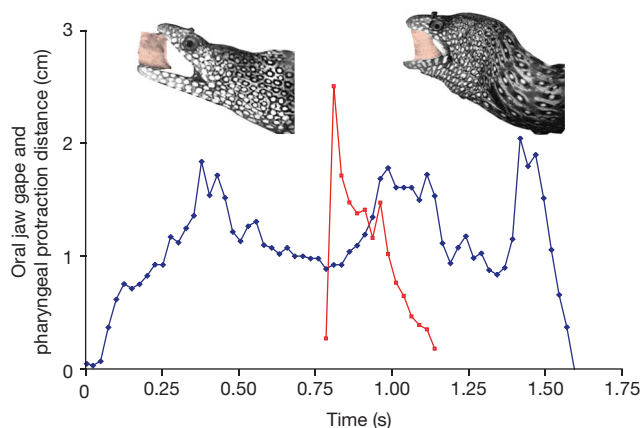


Figure 2 | The oral gape cycle in relation to the pharyngeal jaw cycle in *M. retifera*. After peak oral gape (blue trace), oral jaws make contact with the prey (pink rectangle) by biting. Pharyngeal jaws (red trace) are fully protracted and the recurved teeth on the upper pharyngeal teeth are in contact with the prey. Pharyngeal jaws grip prey and begin retracting prey towards the oesophagus. Prey is pulled into the oesophagus as the moray extends its neurocranium forward and advances its body over the prey while increasing its oral gape, similar to a snake.

us to provide a description of pharyngeal jaw protraction and retraction (Fig. 4a–c). Upper pharyngeal jaw protraction is presumably caused by contraction of the levator externi and interni muscles, which originate on the parasphenoid. The levator externus inserts on the dorsal side of the fourth epibranchial arm. The levator internus muscle inserts on the dorsal side of the upper pharyngobranchial. Contraction will result in dorsal rotation of the pharyngobranchial, which positions the recurved teeth to snag prey on contact. The lower pharyngeal jaws are protracted by the rectus communis, which originates on the ventral side of the hyoid arch and attaches onto the anteroventral margin of the lower pharyngeal jaw. Throughout pharyngeal protraction, the epibranchial and ceratobranchial arms are adducted by contraction of a well-developed adductor muscle that originates on the dorsal epibranchial and inserts onto the lateral side of the ceratobranchial. Adduction of the pharyngeal arms compresses the pharyngeal jaw apparatus, facilitating smooth movement through the pharynx. As the pharyngeal jaw enters the oral cavity the distance between the upper and lower pharyngeal teeth widens owing to further protraction of the upper and lower jaw and relaxation of the adductor.

Once the teeth of the upper jaw ensnare the prey, a dorsal retractor and the pharyngocleithralis retract the pharyngeal jaw. The dorsal retractor originates on the vertebral column (around vertebrae 15 in *M. retifera*) and inserts on the most posterior end of the epibranchial arm. The pharyngocleithralis originates on the cleithrum and inserts onto a lateral groove on the lower pharyngeal

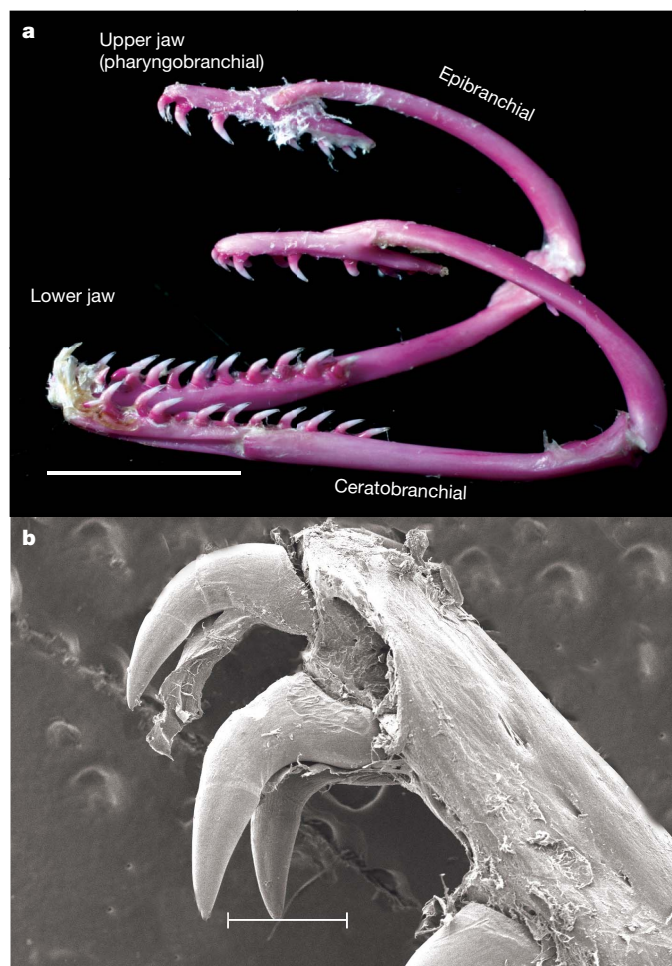


Figure 3 | Detailed anatomy of the pharyngeal jaw apparatus in *M. retifera*. a, Left lateral view of a cleared and alizarin red-stained pharyngeal jaw apparatus, illustrating the sharp, recurved teeth on the pharyngobranchials used to grasp prey. Scale bar, 1 cm. b, Left anterior upper pharyngobranchial revealing highly recurved teeth. Scale bar, 500 μ m.

jaw. The recurved teeth on the lower pharyngeal jaw ensnare prey during retraction rather than protraction. During retraction, the pharyngeal arms adduct so that the upper and lower teeth securely bite down on the prey. With a firm grip on the prey, the pharyngeal jaws travel back to their resting position behind the skull. The prey is further transported into the oesophagus by contraction of the oesophageal sphincter followed by bilateral compression of the body, resulting in posterior-directed waves. Whereas *M. retifera* transported short squid pieces using a single oral-pharyngeal protraction/retraction cycle, long prey (roughly 5 cm in length) were transported using 3–5 cycles (see Supplementary Movie 3). Once the prey had completely entered the oral cavity and could not be grasped by the oral jaws, only pharyngeal and cervical vertebral movements were used to swallow prey.

There are advantages to using a mechanical transport system rather than a hydraulic mechanism to pull prey into the oesophagus. Morays hunt in rocky crevices^{4,7,10,11} and these confined spaces may limit the cranial expansion required to generate intra-oral water movement. Both suction feeding and hydraulic transport mechanisms require rapid rotation and abduction of many cranial elements^{20,24}. The angular excursion of cranial movements scales with body size, and maximum excursion velocities and overall timing of mouth opening increases during suction feeding²⁵. Thus, cranial movements may be limited and less effective for large predatory fish hunting in the confines of coral crevices. Small prey are easily captured with suction, whereas large prey may escape the flow field in

front of a moray's mouth²⁶. Large morays with their well-developed jaw-closing muscles and sharp teeth can potentially generate much greater forces with a bite than with hydrodynamic forces²⁷. Also, teeth in the oral jaws can restrain the prey even if only in contact with a small portion of the prey. As long as the moray can sink a few teeth into its prey, the pharyngeal jaws can protract to deliver a second bite.

Although the 45,000 species of vertebrates exhibit great diversity in feeding mechanisms, very few transport behaviours exist. Hydraulic, lingual and inertial transport behaviours are widespread and have evolved independently numerous times across vertebrate lineages²⁸. A fourth mechanism, ratcheting (gnathic transport), occurs in snakes and exhibits remarkable similarities to moray transport behaviour in which the oral and pharyngeal jaws alternate to pull prey into the oesophagus. The independent evolution of ratcheting mechanisms enables both morays and snakes to maintain a constant grip on their prey. The ability of morays and snakes to circumvent gape constraints that are magnified by their convergent elongate and limbless body plan may be independently correlated with analogous innovations of their feeding apparatus.

Pharyngeal jaws are a complex musculoskeletal system and a major innovation that increases the range of trophic specializations for feeding in teleosts^{2,29}. The evolution of the newly discovered function and design of this widespread jaw system adds to our general understanding of how innovations arise and how they correlate with a particular body plan. Our discovery demonstrates that striking functional novelties can arise with only subtle modifications in existing systems, and offers new insights into the functional morphology of a successful radiation of predatory fish, the moray eels.

METHODS SUMMARY

Individuals of *M. retifera* were filmed in 100-litre aquaria at the University of California at Davis (UCD) using a NAC Memrecam ci digital system. Videofluoroscopy was performed on a single specimen at the Center for Imaging Sciences at the School of Veterinary Medicine, UCD. After feeding trials were obtained, freshly killed specimens were radiographed; they were then fixed in 10% formalin and stored in 70% ethanol for dissections, whole-mount clearing and double-staining³⁰. The tooth morphology of the pharyngeal jaws was examined with a scanning electron microscope at the Electron Microscopy Laboratory, Department of Medical Pathology and Laboratory Medicine, School of Medicine, UCD.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 March; accepted 3 July 2007.

1. Lauder, G. V. in *Fish Biomechanics* (eds Webb, P. W. & Weihs, D.) 280–311 (Praeger, New York, 1983).
2. Wainwright, P. C. in *Fish Biomechanics* (eds Shadwick, R. & Lauder, G. V.) 77–101 (Academic, New York, 2006).
3. Mehta, R. M. & Wainwright, P. C. Biting releases constraints on moray eel feeding kinematics. *J. Exp. Biol.* **210**, 495–505 (2007).
4. Randall, J. E. Food habits of reef fishes of the West Indies. *Stud. Trop. Oceanogr.* **5**, 655–847 (1967).
5. Young, R. F. & Winn, H. E. Activity patterns, diet, and shelter site use for two species of moray eels, *Gymnothorax moringa* and *Gymnothorax vicinus*, in Belize. *Copeia* **2003**, 44–55 (2003).
6. Yukihiro, H., Shibuno, T., Hashimoto, H. & Gushima, K. Feeding habits of moray eels (Pisces: Muraenidae) at Kuchierabu-ji. *Appl. Biol. Sci.* **33**, 159–166 (1994).
7. Parrish, J. D. et al. Piscivory in a coral reef fish community. *Environ. Biol. Fish.* **14**, 285–297 (1986).
8. Nelson, G. J. Branchial muscles in representatives of five eel families. *Pacif. Sci.* **21**, 348–363 (1967).
9. Nelson, G. J. Gill arches of Teleostean fishes of the Order Anguilliformes. *Pacif. Sci.* **20**, 391–408 (1966).
10. Abrams, R. W., Abrams, M. D. & Schein, M. W. Diurnal observations on the behavioral ecology of *Gymnothorax moringa* (Cuvier) and *Muraena miliaris* (Kaup) on a Caribbean coral reef. *Coral Reefs* **1**, 185–192 (1983).
11. Böhlke, E. B., McCosker, J. E. & Böhlke, J. E. in *Fishes of the Western North Atlantic* (ed. Böhlke, E. B.) 104–206 (Sears Foundation for Marine Research Memoir, Yale Univ., New Haven, Connecticut, 1989).
12. Cundall, D. & Greene, H. W. in *Feeding: Form, Function, and Evolution in Tetrapod Vertebrates* (ed. Schwenk, K.) 293–333 (Academic, San Diego, 2000).

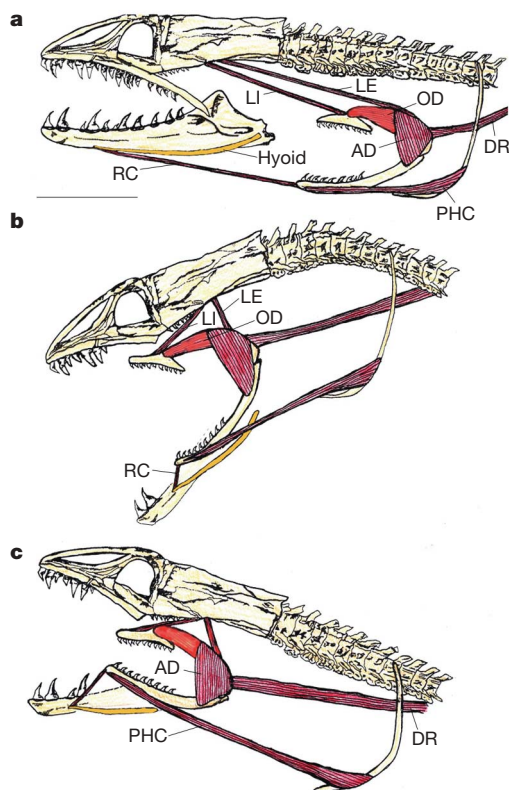


Figure 4 | Functional morphological model of pharyngeal jaw movement in *M. retifera*. The left dentary has been removed in **a–c**, and the left maxilla has been removed in **b** and **c**. **a**, Pharyngeal jaw apparatus at rest. **b**, Pharyngeal jaw protracted: the levator internus (LI) and levator externus (LE) protract the upper jaw into the oral cavity, whereas the rectus communis (RC) protracts the lower jaw. During protraction, the upper pharyngobranchial is dorsally rotated by contraction of the LI and the obliquus dorsalis (OD). **c**, After prey contact, the adductor (AD) contracts to bring the upper and lower jaws together to deliver a second bite. The dorsal retractor (DR) and pharyngocleithralis (PHC) retract the pharyngeal jaws back to their resting position behind the skull. Scale bar, 1 cm.

13. Kley, N. J. & Brainerd, E. L. Post-cranial prey transport mechanisms in the black pinesnake, *Pituophis melanoleucus lodingi*: an x-ray videographic study. *Zoology* **105**, 153–164 (2002).
14. Miller, T. J. Knotting: a previously undescribed feeding behaviour in Muraenid eels. *Copeia* **1987**, 1055–1057 (1987).
15. Robins, C. R. in *Fishes of the Western North Atlantic* (ed. Böhlke, E. B.) 9–23 (Sears Foundation for Marine Research Memoir, Yale Univ., New Haven, Connecticut, 1989).
16. Gilbert, M., Rasmussen, J. B. & Kramer, D. L. Estimating the density and biomass of moray eels (Muraenidae) using a modified visual census method for hole-dwelling reef fauna. *Environ. Biol. Fish.* **73**, 415–426 (2005).
17. Hixon, M. A. & Beets, J. P. Predation, prey refuges, and the structure of coral-reef fish assemblages. *Ecol. Monogr.* **63**, 77–101 (1993).
18. Carr, M. H. & Hixon, M. A. Predation effects on early post-settlement survivorship of coral-reef fishes. *Mar. Ecol. Prog. Ser.* **124**, 31–42 (1995).
19. Alfaro, M. E., Janovetz, J. & Westneat, M. W. Motor control across trophic strategies: muscle activity of biting and suction feeding fishes. *Am. Zool.* **41**, 1266–1279 (2001).
20. Gillis, G. B. & Lauder, G. V. Kinematics of feeding in bluegill sunfish: is there a general distinction between aquatic capture and transport behaviors? *J. Exp. Biol.* **198**, 709–720 (1995).
21. Popta, C. M.-L. Les arcs branchiaux de quelques Muraenidae. *Ann. Sci. Nat.* **19**, 367–390 (1909).
22. Hulsey, C. D., Fraser, G. J. & Streelman, J. T. Evolution and development of complex biomechanical systems: 300 million years of fish jaws. *Zebrafish* **2**, 243–257 (2005).
23. Lauder, G. V. Functional design and evolution of the pharyngeal jaw apparatus in euteleostean fishes. *Zool. J. Linn. Soc.* **77**, 1–38 (1983).
24. Carroll, A. M., Wainwright, P. C., Huskey, S. H., Collar, D. C. & Turingan, R. G. Morphology predicts suction feeding performance in centrarchid fishes. *J. Exp. Biol.* **207**, 3873–3881 (2004).
25. Richard, B. A. & Wainwright, P. C. Scaling of the feeding mechanism of largemouth bass (*Micropterus salmoides*): kinematics of prey capture. *J. Exp. Biol.* **198**, 419–433 (1995).
26. Day, S. W., Higham, T. E., Cheer, A. Y. & Wainwright, P. C. Spatial and temporal flow patterns during suction feeding of bluegill sunfish (*Lepomis macrochirus*) by Particle Image Velocimetry. *J. Exp. Biol.* **208**, 2661–2671 (2005).
27. Westneat, M. W. Evolution of levers and linkages in the feeding mechanism of fishes. *Integr. Comp. Biol.* **44**, 378–389 (2004).
28. Schwenk, K. in *Feeding: Form, Function and Evolution in Tetrapod Vertebrates* (ed., Schwenk, K.) 21–61 (Academic, San Diego, 2000).
29. Liem, K. F. Evolutionary strategies and morphological innovations: cichlid pharyngeal jaws. *Sys. Zool.* **22**, 425–441 (1973).
30. Dingerkus, G. & Uhler, L. D. Enzyme clearing of alcian blue stained whole small vertebrates for demonstration of cartilage. *Stain Technol.* **52**, 229–232 (1977).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. E. Alfaro, D. C. Collar, S. W. Day, H. W. Greene, N. J. Kley, A. H. Krakauer, R. D. Mehta, J. T. Redwine, A. Sinsheimer, T. W. Schoener and K. Vanderveen for comments on various drafts of this manuscript. We thank N. J. Kley for discussions regarding prey transport in vertebrates. H. Tran and L. B. Feng provided assistance with video collections. We are grateful to R. E. Pollard, C. Stafford, T. B. Waltzek and E. R. Wisner for help with radiographs and videofluoroscopy. P. Kysar provided technical assistance with scanning electron microscopy. We thank the California Academy of Sciences and the Harvard Museum of Comparative Zoology for specimen loans. Funding was provided by the American Association of University Women and by the National Science Foundation.

Author Contributions R.S.M. designed research, performed research and wrote the manuscript. P.C.W. helped design the study, contributed to the interpretation of anatomy and video, and participated in writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.S.M. (rsmehta@ucdavis.edu).

METHODS

Five adult *M. retifera* (standard lengths 35.5, 37.2, 34.2, 36.6 and 40.3 cm) were obtained commercially. Individuals of *M. retifera* were filmed feeding on pieces of cut squid (*Loligo* sp.). Individuals were housed and filmed at 22–27 °C using a NAC Memrecam ci digital system with illumination from two 600-W flood lights. Video sequences were recorded at 100 images per second. Distances in the images were scaled by recording an image of a ruler placed in the field of view.

After all feeding sequences were obtained, pharyngeal jaw movements were studied using a radiograph/digital fluoroscope (Philips, Omni diagnost Eleva) at the Center for Imaging Sciences at the School of Veterinary Medicine, UCD. A moray was placed in a 40-litre glass aquarium and was filmed at 60 images s⁻¹ while consuming a single goldfish, *Carassius auratus*, soaked in 60% w/v liquid barium sulphate (Novopaque, LPI Diagnostics). Afterwards, the same moray was killed using tricaine methane sulphonate (MS-222) and immediately prepared for radiographs. Radiographs were taken to show the pharyngeal jaws in lateral view at rest and in the extreme protracted position. To radiograph the pharyngeal jaw in the extreme protracted position, we protracted the pharyngeal jaws of the specimen with forceps. The distance the pharyngeal jaw was protracted was slightly less than we observed in the video sequences.

The remaining specimens were killed by overexposure to MS-222 and were formalin-fixed to examine the anatomy related to the pharyngeal jaw apparatus. All specimens were fixed in buffered 10% formalin and stored in 70% ethanol. Anatomical dissections were performed on three specimens to understand the musculature used for protracting and retracting the pharyngeal jaws. Two specimens were cleared using trypsin and double-stained in alcian blue cartilage stain and alizarin red bone stain³⁰. Following staining, a pharyngeal jaw from a single specimen was removed and cleaned using forceps and warm water. We then used a digital Canon EOS camera with a macro lens to take photographs of the pharyngeal jaw.

Tooth morphology of the pharyngeal jaws was examined with a scanning electron microscope (Philips XL30 TMP, FEI Co.) and examined with iTEM Software. Our sample was rinsed in deionized water before dehydration in a graded ethanol series, polished with gold powder, dried and then mounted on aluminium stubs and sputter-coated with gold–palladium.

Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution

Debbie Lindell^{1†}, Jacob D. Jaffe^{3†}, Maureen L. Coleman¹, Matthias E. Futschik⁵, Ilka M. Axmann⁵, Trent Rector⁴, Gregory Kettler¹, Matthew B. Sullivan¹, Robert Steen⁴, Wolfgang R. Hess⁶, George M. Church³ & Sallie W. Chisholm^{1,2}

Interactions between bacterial hosts and their viruses (phages) lead to reciprocal genome evolution through a dynamic co-evolutionary process^{1–5}. Phage-mediated transfer of host genes—often located in genome islands—has had a major impact on microbial evolution^{1,4,6}. Furthermore, phage genomes have clearly been shaped by the acquisition of genes from their hosts^{2,3,5}. Here we investigate whole-genome expression of a host and phage, the marine cyanobacterium *Prochlorococcus* MED4 and the T7-like cyanophage P-SSP7, during lytic infection, to gain insight into these co-evolutionary processes. Although most of the phage genome was linearly transcribed over the course of infection, four phage-encoded bacterial metabolism genes formed part of the same expression cluster, even though they are physically separated on the genome. These genes—encoding photosystem II D1 (*psbA*), high-light inducible protein (*hli*), transaldolase (*talC*) and ribonucleotide reductase (*nrd*)—are transcribed together with phage DNA replication genes and seem to make up a functional unit involved in energy and deoxynucleotide production for phage replication in resource-poor oceans. Also unique to this system was the upregulation of numerous genes in the host during infection. These may be host stress response genes and/or genes induced by the phage. Many of these host genes are located in genome islands and have homologues in cyanophage genomes. We hypothesize that phage have evolved to use upregulated host genes, leading to their stable incorporation into phage genomes and their subsequent transfer back to hosts in genome islands. Thus activation of host genes during infection may be directing the co-evolution of gene content in both host and phage genomes.

Prochlorococcus is the dominant photosynthetic organism in vast regions of the world's oceans⁷, where T7-like podoviruses are also abundant⁸. Therefore this phage–host system is likely to be of great relevance for bacterial and phage global evolution, for modelling their population dynamics, and for understanding the role of phage in the oceanic carbon cycle.

Phages infecting marine cyanobacteria encode a number of host-like genes including photosynthesis and stress-response genes^{5,9–11}. Phage photosynthesis genes are expressed during infection while transcripts of homologous genes in the host decline^{12,13}, and are hypothesized to facilitate production of carbon and energy through cell photosynthesis for optimal phage production^{5,10,12–14}. This physiological interdependence between host and phage is likely to have led to the observed prevalence of photosynthesis genes in cyanophage^{10,15}, providing a reservoir for genetic exchange, and influencing the co-evolutionary process of both host and phage^{14,15}.

Although the analysis of single genes has provided insight into this dynamic, a systems approach is essential for a broader understanding of this co-evolutionary process. Here we investigate genome-wide transcriptome dynamics of *Prochlorococcus* MED4 and the T7-like podovirus P-SSP7 over the course of infection—the first such detailed view of infection for any lytic host–phage system.

We first characterized the gross features of the lytic cycle (Fig. 1). Phage genomic DNA (gDNA) began to increase, 4 h after infection, and host gDNA to decrease, 4 h after infection, and phage progeny were first released into the extracellular medium 8 h post infection (Fig. 1a). Phage

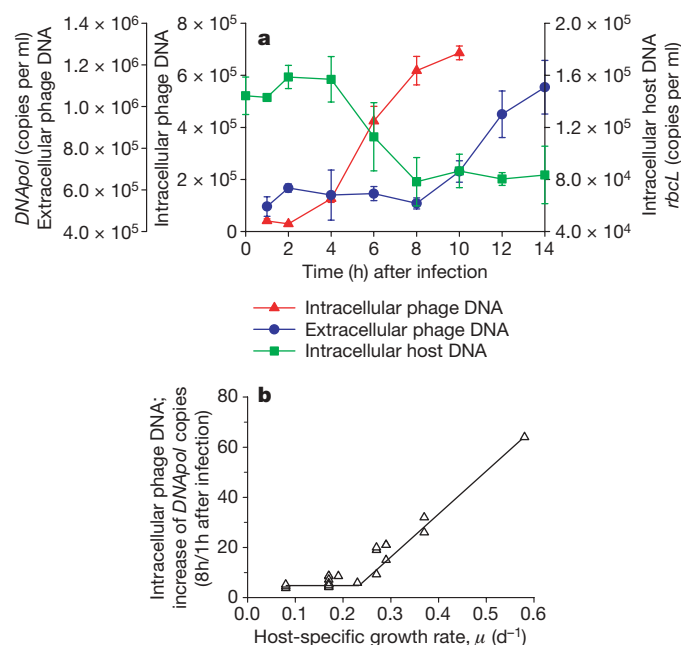


Figure 1 | Infection dynamics of *Prochlorococcus* MED4 by podovirus P-SSP7. a, Timing of phage gDNA replication (intracellular phage DNA) and length of the lytic cycle (extracellular phage DNA) was determined by quantifying the phage DNA polymerase gene (*gene 5/DNAPol* gene copy number). Host gDNA degradation (intracellular host DNA) was determined by disappearance of the host *rbcl* gene. Average and s.d. of three biological replicates. **b**, Dependence of phage gDNA replication on host growth rate. Phage DNAPol intracellular copy number was measured 8 h after infection and normalized to that at 1 h after infection as a measure for phage gDNA replication. $n = 24$.

¹Department of Civil and Environmental Engineering, ²Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Department of Genetics, and ⁴BioPolymers Facility, Department of Genetics Harvard Medical School, Boston, Massachusetts 02115, USA. ⁵Institute for Theoretical Biology, Humboldt University, Berlin 10115, Germany. ⁶Institute of Biology, University of Freiburg, Freiburg 79104, Germany. [†]Present addresses: Department of Biology, Technion—Israel Institute of Technology, Haifa 32000, Israel (D.L.); The Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02141, USA (J.D.J.).

replication is a function of host growth rate (Fig. 1b), which, together with a dependence on photosynthesis¹², suggests an intimate link between phage fitness and host physiology. We wanted to know: what are the dynamics of the phage and host transcriptome during infection, and where do phage-encoded 'bacterial-like' genes fit into this transcriptional program?

The genome content and architecture of the cyanophage P-SSP7 are similar to that of the *Escherichia coli*-infecting T7 podovirus¹¹. As in T7 (ref. 16), the P-SSP7 genome was transcribed linearly from the left to the right of the genome map over the course of infection (with important exceptions, see below) (Fig. 2), and three expression clusters were discerned (Fig. 2a, and Supplementary Fig. 1). The first cluster contains a putative *marR* transcriptional regulator gene, a T7 homologue (g0.7) suggesting a role in redirecting transcription from the host towards the phage. The second cluster contains genes involved in DNA metabolism and replication (Fig. 2a, and Supplementary Fig. 1) as well as RNA polymerase (RNAP), which may be involved in RNA transcription and/or DNA replication¹⁶. The third cluster consists of genes involved in phage particle formation and DNA maturation. Proteins encoded by this latter cluster were detected in the mature phage particle (Fig. 2b, and Supplementary Table 1), further supporting that many are phage structural genes. Thus the three expression clusters in this cyanophage are analogous to T7-coliphage class I, II and III genes in both gene content¹¹ and the timing of genome expression (Fig. 2). That these fundamental operational properties are conserved across cyanophage and enteric phage, the hosts of which are drastically different with respect to energy

source (autotroph versus heterotroph), habitat (nutrient-poor oceanic waters versus the nutrient-rich human gut), and growth rate (generation time of a day versus less than an hour), is remarkable.

Despite the similar overall infection strategies of P-SSP7 and T7, transcription cluster 2 in the cyanophage displays novel features in both gene content and regulation and bears signatures of host-phage co-evolution unique to the marine ecosystem. This cluster contains four 'bacterial-like' genes: the ribonucleotide reductase gene *nrd* (ORF 020), the high-light-inducible stress response gene *hli* (ORF 026), the photosystem II gene *psbA* (ORF 027), and the transaldolase gene *talC* (ORF 054). Although *nrd*, *hli* and *psbA* are in the middle of the genome, *talC* is at the end¹¹ (Fig. 2b). The co-transcription of these four genes, despite their physical separation (Fig. 2, and Supplementary Fig. 2), suggests that they are functionally linked³.

Clues as to the function of the 'bacterial-like' genes are given by their position in the transcriptional and translational program of the entire host-phage system. First, the proteins encoded by these genes are present during infection but absent from the mature phage particle (Fig. 2b, and Supplementary Table 1), indicating that they function intracellularly. Second, these genes are transcribed together with DNA replication genes, and include ribonucleotide reductase, which converts host ribonucleotides, recycled from degraded RNA (see below), to deoxynucleotides. The photosynthesis genes found in this cluster are thought to be involved in the production of energy^{5,10,12–14} and transaldolase may function in the host's pentose phosphate pathway to produce reducing power (NADPH) and/or ribose substrates for nucleotide synthesis¹¹. Together, these findings suggest that these genes form a functional unit to produce energy and deoxynucleotide carbon substrates necessary for cyanophage DNA replication in the resource-poor oceans.

The bacterial-like metabolism genes found in P-SSP7 are also commonly found in myoviruses that infect marine cyanobacteria, despite drastic differences in their core genome content^{9,11}. In some myoviruses, however, the genes are situated together on the genome^{10,11}. Therefore we may be seeing a snapshot of evolution in progress, from spatial separation with cotranscription in P-SSP7, to physically linked genes in other cyanophage genomes.

It is not at all clear how the transcription of this cyanophage genome is regulated, and, in particular, how the last three genes are co-regulated with cluster 2 genes. Although we bioinformatically detected host-like RNAP promoters upstream of each phage expression cluster, and ORF 052, no clear phage-like RNAP promoters were detected¹⁷ (Supplementary Table 2). A transcription initiation site consistent with bacterial-like promoters was experimentally mapped upstream of cluster 2 genes, whereas 5' ends consistent with RNA processing sites and with weak similarity to T7-like promoters, were found upstream of cluster 1 and cluster 3 genes (Supplementary Fig. 3). However, it remains unclear whether these sequences serve as phage promoters and/or RNA processing sites for transcripts generated by either host or phage RNAP.

Given the reliance of phage replication on host physiology (Fig. 1b), the behaviour of the host transcriptome during infection is of interest. Whereas the transcript levels for approximately 75% of the 1,716 host genes declined during infection (Fig. 3), 41 genes were significantly upregulated. This is distinctly different from other lytic host-phage systems where few, if any, host genes become activated^{16,18}. The upregulated genes fall into two groups (Fig. 3, and Supplementary Fig. 4 and Supplementary Table 3). The first was transiently upregulated immediately after infection and consists of high-light-inducible stress response (*hli*), carbon metabolism (*rbclS*), transcription (*rpoC2*, *rpoD*) and ribosome (*rpl5*, *rpl6*, *rps8*, *rps11*, *rps17*) genes. Transcripts of the second group appeared 2 h after infection and included genes involved in RNA degradation and modification (*rne*, *rnhB*, *dus* and *sun*), protein turnover (*clpS*, and an AAA ATPase family gene), stress responses (*umuD* and *phoH*), and those of unknown function. Two of the latter were

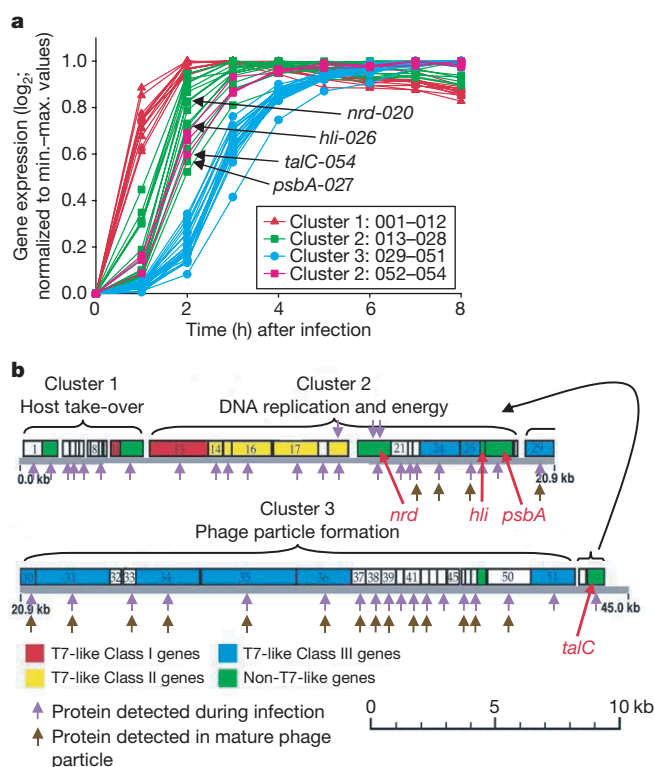


Figure 2 | Temporal expression dynamics of P-SSP7 phage genes during infection of *Prochlorococcus* MED4. **a**, Transcript levels with time after infection reveal three transcription clusters (see Supplementary Fig. 1). Profiles determined from microarrays, were normalized to minimum–maximum values for each gene. Average of three biological replicates; Supplementary Fig. 6 shows RT–PCR verification of results. **b**, Genome map¹¹ highlighting the position of *talC* at the end of the genome, even though it is transcribed in cluster 2. Protein detection during infection (purple arrows) and in mature phage particle (brown arrows) showing that 74% of phage genes produced proteins, including three overlapping genes that escaped previous annotation (Supplementary Table 6). Supplementary Table 1 encompasses gene identification and peptide detection.

transcribed from bacterial-like promoters (Supplementary Fig. 3), suggesting involvement of host RNAP. Upregulated host gene expression may constitute a direct stress response to phage infection, or may have been facilitated by phage factors¹⁶ injected into the cell or expressed from phage expression cluster 1.

Regardless of the mechanism of upregulation, we hypothesize that phages may have evolved to make use of the products of some upregulated host genes as part of the 'arms-race' between host and phage¹⁹. Certainly, phages are known to exploit host stress-response proteins during infection in other systems^{20–22}. The T4 and T7 phages infecting *E. coli*, for example, have evolved to modify host RNase E (involved in RNA degradation) leading to the degradation of host RNA²². It is perhaps not coincidental that *rne* (encoding RNase E) is one of the upregulated genes in *Prochlorococcus* during infection. This may have initially served as a host defence mechanism for degrading phage RNA, but could also be exploited by phage to degrade host RNA for use as substrates for phage deoxynucleotide synthesis.

Perhaps the most compelling evidence that upregulated host genes are part of the co-evolutionary process in this system is that 34% of them (more than would occur by chance $P < 0.001$) are found in hypervariable host genome islands (Supplementary Table 3), which are thought to be mobilized by phages⁶. Furthermore, homologues of a number of these host genes are found in phage genomes, including *hli*, *phoH*, and HNH endonuclease and sigma factor genes, as well as RNase H and heat-shock genes^{9,11}.

Thus there seems to be a connection between genes upregulated during infection, their position in the host genome, and the presence of homologues in phage. Although there are a number of possible explanations for this connection, the most parsimonious evolutionary scenario is as follows: Host stress response genes are upregulated in response to phage infection. Phages that have evolved to use these gene products gain a fitness advantage. Random incorporation³ of these genes into their own genomes would enable phages to more

tightly regulate their expression, conferring a fitness advantage, and leading to preferential retention. This retention would increase the probability of transfer back to the host in genome islands, by lysogeny or unsuccessful infection, and those genes beneficial to the host would remain in the host genome. Analysis of the *hli* gene family provides an interesting illustration in support of this scenario. *hli* genes are upregulated in the host in response to phage infection (Fig. 3, and Supplementary Table 3), are common in *Prochlorococcus*-infecting phage genomes^{5,11}, and multiple phage-like copies⁵ are found in *Prochlorococcus* genome islands⁶ (Supplementary Table 4). Furthermore, their differential expression in *Prochlorococcus* in response to various environmental stressors (Supplementary Table 4) and the presence of a binding site for the nitrogen transcriptional regulator NtcA upstream of the nitrogen-regulated *hli10* gene²³, suggests that copies acquired from phage⁵ have undergone specialization of function in the host. It remains to be seen whether host fitness has been enhanced by the acquisition of these *hli* genes from cyanophages.

This system-wide analysis of the infection of a cyanobacterium by a phage has led to new insights and hypotheses regarding co-evolutionary interactions between host and phage. These interactions clearly shape the gene content of both host and phage, and probably play a role in shaping the distribution and abundance of cyanobacterial ecotypes in the oceans.

METHODS SUMMARY

Prochlorococcus MED4 was grown at 21 °C under 10–25 $\mu\text{mol photon m}^{-1} \text{s}^{-1}$ continuous white light in Pro99 seawater medium with HEPES and sodium bicarbonate. The length of the lytic cycle was determined by quantifying phage DNA in the extracellular medium using a real-time quantitative PCR (qPCR) assay (see Supplementary Fig. 5 for a comparison with standard methods). The timing of phage DNA replication and host DNA degradation were determined intracellularly using qPCR assays for the phage *DNApol* and host *rbcl* genes, respectively. For expression analysis triplicate cultures (10^8 cells ml^{-1}) were infected with the P-SSP7 podovirus (3×10^8 infective phage particles ml^{-1}) and the paired control cultures were amended with filter-sterilized spent medium. Samples were collected by centrifugation, resuspended in storage buffer and snap frozen in liquid nitrogen. RNA was extracted using Ambion's *mirVana* RNA isolation kit and DNA was removed by DNaseI digestion using Ambion's Turbo DNA-free kit. Transcriptional analyses were carried out using a custom-made high-density antisense Affymetrix array—MD4-9313. Two micrograms of total RNA were subjected to Affymetrix protocols for *E. coli*. Array analyses were carried out using R and Bioconductor, and array data were normalized and probe set summaries calculated using the robust multi-array average (RMA) procedure²⁴. Array results were validated by RT-PCR (Supplementary Figs 6, 7) and the appropriate normalization method was determined by comparing normalized transcription profiles to RT-PCR results (Supplementary Table 5 and Supplementary Figs 8, 9, 10). Promoters were computationally predicted and experimentally assessed using the 5' RACE technique. For the detection of phage proteins, *Prochlorococcus* cells were harvested 3 and 7 h after infection with phage, and 10^{10} caesium-chloride-purified phage particles were subjected to mass spectrometry proteomic analysis as in ref. 25. See Supplementary Methods for details of all experimental procedures.

Received 5 June; accepted 26 July 2007.

1. Canchaya, C., Proux, C., Fournous, G., Bruttin, A. & Brussow, H. Prophage genomics. *Microbiol. Mol. Biol. Rev.* **67**, 238–276 (2003).
2. Filee, J., Forterre, P. & Laurent, J. The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. *Res. Microbiol.* **154**, 237–243 (2003).
3. Hendrix, R. W., Lawrence, J. G., Hatfull, G. F. & Casjens, S. The origins and ongoing evolution of viruses. *Trends Microbiol.* **8**, 504–508 (2000).
4. Hsiao, W. W. L. et al. Evidence of a large novel gene pool associated with prokaryotic genome islands. *PLoS Genet.* **1**, e62 (2006).
5. Lindell, D. et al. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl Acad. Sci. USA* **101**, 11013–11018 (2004).
6. Coleman, M. L. et al. Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**, 1768–1770 (2006).
7. Partensky, F., Hess, W. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63**, 106–127 (1999).
8. Breitbart, M., Miyake, J. H. & Rohwer, F. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol. Lett.* **236**, 249–256 (2004).

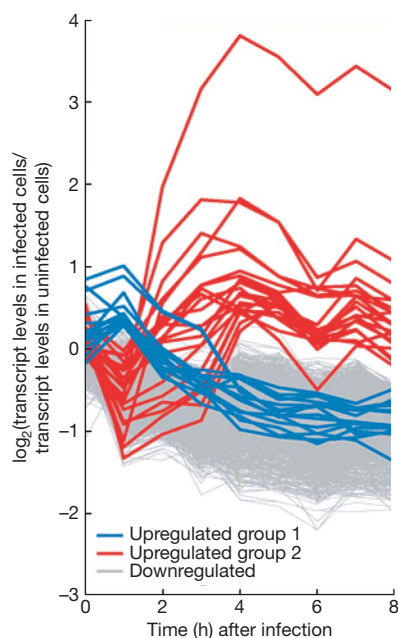


Figure 3 | Transcriptional profiles of *Prochlorococcus* MED4 genes with time after infection by P-SSP7. Transcript levels, determined from microarrays, are presented as \log_2 -fold change in infected cells relative to uninfected cells over the 8 h latent period of infection. Only genes whose expression levels were significant at a false-discovery rate of $q < 0.05$ are shown. Blue and red indicate significantly upregulated genes in transcription groups 1 and 2, respectively (see Supplementary Table 3). Grey indicates genes significantly downregulated at 8 h after infection. Average of three biological replicates. Supplementary Fig. 7 shows RT-PCR verification of results.

9. Mann, N. H. *et al.* The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus*. *J. Bacteriol.* **187**, 3188–3200 (2005).
10. Millard, A., Clokie, M. R., Shub, D. A. & Mann, N. H. Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus* strains. *Proc. Natl Acad. Sci. USA* **101**, 11007–11012 (2004).
11. Sullivan, M. B., Coleman, M., Weigele, P., Rohwer, F. & Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol.* **3**, e144 (2005).
12. Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**, 86–89 (2005).
13. Clokie, M. R. J., Shan, J., Bailey, S., Jia, Y. & Krisch, H. M. Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environ. Microbiol.* **8**, 827–835 (2006).
14. Zeidner, G. *et al.* Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. *Environ. Microbiol.* **7**, 1505–1513 (2005).
15. Sullivan, M. B. *et al.* Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* **4**, e234 (2006).
16. Molineux, I. in *The Bacteriophages* (ed. Calendar, R.) 277–301 (Oxford University Press, New York, 2005).
17. Chen, Z. & Schneider, T. D. Information theory based T7-like promoter models: classification of bacteriophages and differential evolution of promoters and their polymerases. *Nucleic Acids Res.* **33**, 6172–6187 (2005).
18. Miller, E. S. *et al.* Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* **67**, 86–156 (2003).
19. Lenski, R. E. & Levin, B. R. Constraints on the coevolution of bacteria and virulent phage: A model, some experiments, and predictions for natural communities. *Am. Nat.* **124**, 585–602 (1985).
20. Tabor, S., Huber, H. E. & Richardson, C. C. *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *J. Biol. Chem.* **262**, 16212–16223 (1987).
21. Tilly, K., Murialdo, H. & Georgopoulos, C. P. Identification of a second *Escherichia coli* *groE* gene whose product is necessary for bacteriophage morphogenesis. *Proc. Natl Acad. Sci. USA* **78**, 1629–1633 (1981).
22. Ueno, H. & Yonesaki, T. Phage-induced change in the stability of mRNAs. *Virology* **329**, 134–141 (2004).
23. Tolonen, A. C. *et al.* Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. *Mol. Systems Biol.* **2**, 53 (2006).
24. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
25. Jaffe, J. D. *et al.* The complete genome and proteome of *Mycoplasma mobile*. *Genome Res.* **14**, 1447–1461 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Steglich, S. Bhattacharya, H. Keller, L. Thompson, P. Weigele, S. Choe, D. Endy, S. Kosuri, M. Shmoish and J. Aach for discussions, and J. Waldbauer and M. Osburne for comments on the manuscript, and the MIT Center for Environmental Health Sciences. This work was funded by the DOE Genomes to Life System Biology Center Grant (G.M.C. and S.W.C.), the Gordon and Betty Moore Foundation's Marine Microbiology Program (S.W.C.), and the National Science Foundation (S.W.C.).

Author Information The microarray data have been deposited in the GEO database under the accession number GSE8382. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.W.C. (chisholm@mit.edu).

Cadherin 23 and protocadherin 15 interact to form tip-link filaments in sensory hair cells

Piotr Kazmierczak^{1,2*}, Hirofumi Sakaguchi^{3*}, Joshua Tokita³, Elizabeth M. Wilson-Kubalek¹, Ronald A. Milligan¹, Ulrich Müller^{1,2*} & Bechara Kachar^{3*}

Hair cells of the inner ear are mechanosensors that transduce mechanical forces arising from sound waves and head movement into electrochemical signals to provide our sense of hearing and balance. Each hair cell contains at the apical surface a bundle of stereocilia. Mechano-electrical transduction takes place close to the tips of stereocilia in proximity to extracellular tip-link filaments that connect the stereocilia and are thought to gate the mechano-electrical transduction channel^{1–3}. Recent reports on the composition^{4–8}, properties and function^{9–11} of tip links are conflicting²⁹. Here we demonstrate that two cadherins that are linked to inherited forms of deafness in humans^{12–15} interact to form tip links. Immunohistochemical studies using rodent hair cells show that cadherin 23 (CDH23) and protocadherin 15 (PCDH15) localize to the upper and lower part of tip links, respectively. The amino termini of the two cadherins co-localize on tip-link filaments. Biochemical experiments show that CDH23 homodimers interact *in trans* with PCDH15 homodimers to form a filament with structural similarity to tip links. Ions that affect tip-link integrity and a mutation in *PCDH15* that causes a recessive form of deafness¹⁶ disrupt interactions between CDH23 and PCDH15. Our studies define the molecular composition of tip links and provide a conceptual base for exploring the mechanisms of sensory impairment associated with mutations in CDH23 and PCDH15.

Ultrastructural studies show that tip links are helical filaments that separate into multiple strands at both ends^{9,11}. Tip links share properties with cadherins^{17,18}, and CDH23 and PCDH15 have been localized to tip links^{4,7}. However, several laboratories failed to detect CDH23 expression at tip links^{5,6}, and an antibody to the extracellular domain of PCDH15 did not label tip links in mature mammalian hair cells⁴. To clarify the localization of CDH23 and PCDH15 in hair cells, we generated antibodies against CDH23 and PCDH15. Antibodies PB264 and PB240 were raised against peptide sequences corresponding to regions between the extracellular cadherin (EC) domains EC1/EC2 and EC15/EC16 of CDH23, respectively. Antibody PB811 was raised against a peptide sequence in EC1 of PCDH15. In western blots, PB264 and PB240 specifically recognized CDH23, whereas PB811 recognized PCDH15 (Supplementary Fig. 1). In agreement with earlier findings^{4–7}, immunolocalization experiments revealed expression of CDH23 and PCDH15 associated with kinociliary links and transient lateral links in developing hair cells of mice, rats and guinea-pigs (Supplementary Fig. 2). In addition, the three antibodies stained tip links in mature hair cells (Fig. 1a–e; see also Supplementary Fig. 2). CDH23 labelling appeared as fluorescent puncta at the tip-link region (Fig. 1a–e). In splayed stereocilia, the puncta predominantly partitioned with the taller stereocilia (Fig. 1b, c). Removal of Ca²⁺ causes breakage of tip links and a relaxation movement in the

stereociliary bundle¹⁷. Similarly, when samples were pre-incubated for 1 min in 5 mM BAPTA to remove Ca²⁺ and allowed to recover for 10 min at 25 °C or at 37 °C, the CDH23 fluorescent puncta separated with the taller stereocilia (Fig. 1d, e). At 37 °C, the fluorescent puncta appeared closer to the stereocilia tips, suggesting temperature-dependent movement and further separation of the upper portion of the tip link after BAPTA treatment. These findings are consistent with CDH23 localizing to the upper part of tip links.

PCDH15 localizes to the tips of stereocilia⁴ (Supplementary Fig. 2f, g) and may interact with CDH23 to form tip links. We therefore used immunogold transmission electron microscopy (TEM) to analyse the localization of PCDH15 and CDH23. Interactions between classical cadherins such as N- and E-cadherin depend on EC1 (ref. 18). Each EC domain has a dimension of ~4.5 nm¹⁹. Assuming a similar mode of interaction and dimensions for the EC domains in CDH23 and PCDH15, PB811 is expected to bind at ~49.5 nm (11 ECs) from the lower end of tip links, PB264 at ~54 nm (12 ECs) and PB240 at ~117 nm (26 ECs) (Fig. 1j). Using PB811, PB264 and PB240, we observed gold labelling at 37 ± 17 nm (*n* = 113), 52.5 ± 19 nm (*n* = 111) and 138 ± 34 nm (*n* = 52) above the lower tip-link insertion site, respectively (Fig. 1f–i). These values agree with the predicted values and suggest that CDH23 and PCDH15 molecules interacting at their N termini form tip links (Fig. 1k). However, we cannot determine from our gold labelling analysis how many CDH23 and PCDH15 molecules form the tip link, and some gold particles associated with CDH23 were further apart from the lower insertion point of the tip link than predicted, which could be a consequence of tip-link breakage during sample preparation. Because it is difficult to visualize tip links in thin sections, we confirmed by freeze-etching immunogold labelling that the CDH23 antibody stained the tip-link filament (Supplementary Fig. 3).

To determine whether CDH23 and PCDH15 molecules resemble tip links, we expressed, by transfection in HEK293 cells, the extracellular domains of CDH23 and PCDH15 fused to His affinity tags (Fig. 2a). The fusion proteins were purified with nickel (Ni-NTA) beads (Fig. 2b) and analysed by negative staining TEM. In the presence of 1 mM Ca²⁺, CDH23–His molecules formed dimeric filaments with helical appearance (Fig. 2c and Supplementary Fig. 4). Frequently the two strands splayed at one filament end and formed a branched or looped structure. The extracellular domain of CDH23 consists of 27 EC domains with an estimated size of ~122 nm, which is in good agreement with the observed length of 129 ± 4.5 nm. In the absence of Ca²⁺, the CDH23 strands lost their filamentous shape, suggesting Ca²⁺-dependent folding of CDH23 (Supplementary Fig. 4). The extracellular domain of PCDH15 also formed intertwined dimers (Fig. 2e and Supplementary Fig. 5) with a length of

¹The Scripps Research, Institute Department of Cell Biology, ²Institute for Childhood and Neglected Disease, La Jolla, California 92037, USA. ³Laboratory of Cellular Biology, National Institute on Deafness and other Communication Disorders, National Institutes of Health, Bethesda, Maryland 20892, USA.

*These authors contributed equally to this work.

51.9 ± 2.7 nm, which is in agreement with the ~ 49.5 nm predicted for PCDH15.

To define the orientation of extracellular domains of CDH23 and PCDH15 in homodimers, we used Ni-NTA beads coupled to nanogold particles (Fig. 2d, f and Supplementary Fig. 5). Several Ni-NTA beads can bind to each His tag²⁰. For CDH23, we analysed 195 filaments. A total of 131 showed electron densities consistent with nanogold particles at one end, whereas only 3 showed nanogold particles at both ends (Fig. 2d, f and Supplementary Fig. 5). For PCDH15, we analysed 58 filaments. A total of 39 were labelled at one end and the remaining filaments were not labelled. In CDH23, the gold beads attached to the splayed end of the filaments (Fig. 2d), suggesting that the branch point is membrane proximal. These findings suggest that the extracellular domains of CDH23 and PCDH15 form preferentially parallel homodimers.

The localization of CDH23 and PCDH15 at opposite tip-link ends suggests that they interact to form tip links. To analyse interactions, we generated constructs containing the extracellular domain of CDH23 and PCDH15 fused to the Fc domain of the immunoglobulin heavy chain (Fig. 3a). We independently expressed CDH23-His, CDH23-Fc, PCDH15-His and PCDH15-Fc, mixed the proteins in the presence of 1 mM Ca^{2+} , and isolated protein complexes using

Ni-NTA beads, which bind to the His tag but not the Fc tag. Western blot analysis revealed that PCDH15-Fc, but not CDH23-Fc, interacted with CDH23-His (Fig. 3b). PCDH15-His interacted with CDH23-Fc but not with PCDH15-Fc (Fig. 3c). We conclude that CDH23 and PCDH15 engage in heterophilic but not homophilic interactions. We were surprised that CDH23-His and CDH23-Fc (or PCDH15-His and PCDH15-Fc) did not interact because electron microscopic analysis revealed lateral interaction. However, most of the CDH23 (and PCDH15) molecules appeared in electron micrograph images as dimers (Fig. 2 and Supplementary Figs 4 and 5), suggesting that the proteins are purified as stable dimers. Furthermore, whereas CDH23 can mediate homophilic binding⁷, the washing conditions used here were of greater stringency than in previous studies and suggest that homophilic binding is of low affinity. Homophilic adhesion could be important in cells expressing high CDH23 levels.

Adhesion between classical cadherins is optimal at ≥ 1 mM Ca^{2+} (ref. 18), whereas tip links are bathed in the sub-millimolar Ca^{2+} concentration of the endolymph²¹. Interactions between CDH23 and PCDH15 were disrupted in the absence of Ca^{2+} , but were maximal at ≥ 0.1 mM Ca^{2+} , which is within the range of the Ca^{2+} concentration of the endolymph²¹. Tip links are disrupted by treatment with LaCl_3 , but not by treatment with BaCl_2 (ref. 9). Similarly, complex

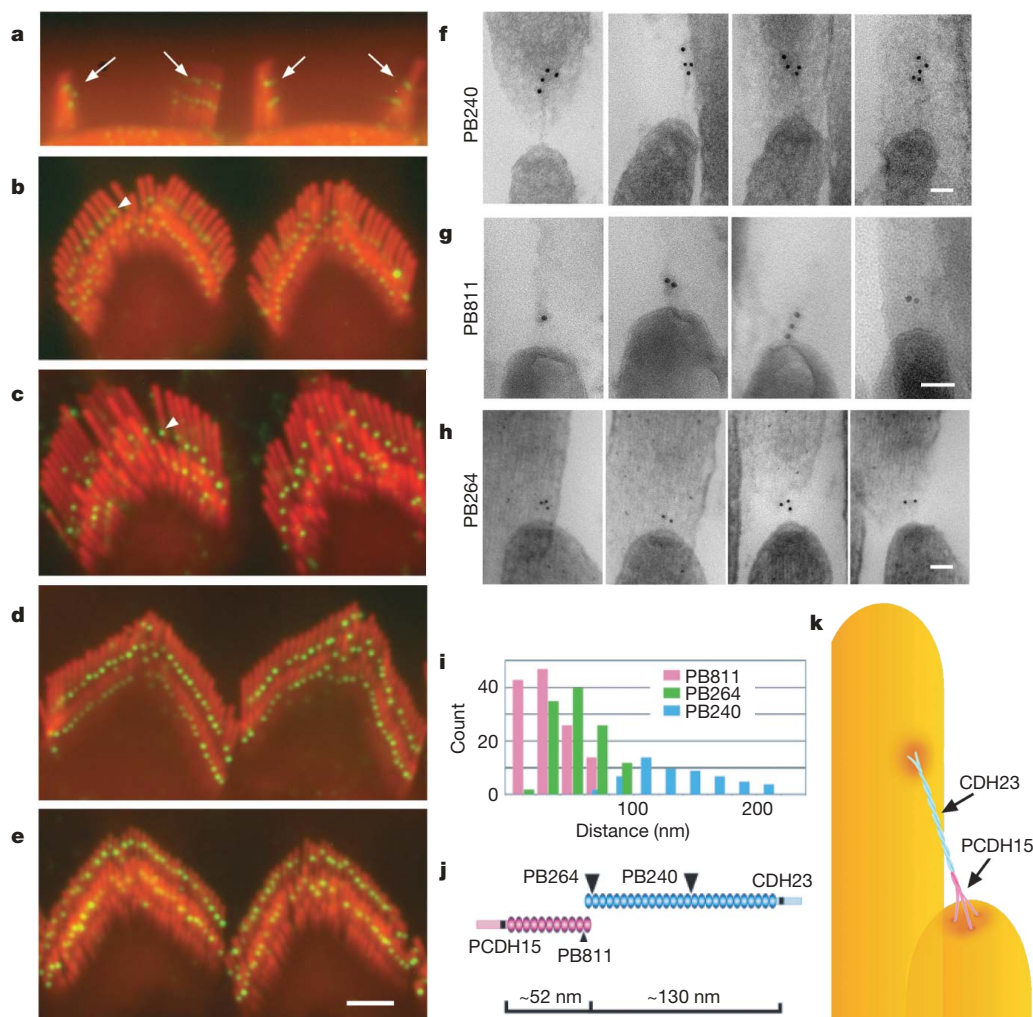


Figure 1 | Tip links are formed by CDH23 and PCDH15. **a–e**, CDH23 antibodies PB240 (**a**, **b**, **d**, **e**) and PB264 (**c**) labelled (green) the tip-link region (arrows) in hair cells of adult guinea-pigs (**a**, **c–e**) and rats (**b**). F-actin was visualized with rhodamine–phalloidin (red). In splayed stereocilia (**b**, **c**) CDH23 fluorescent puncta frequently localized at the lateral wall of stereocilia (arrowheads). After treatment of samples with BAPTA and recovery at 25 °C (**d**) or at 37 °C (**e**), CDH23 separated with taller stereocilia.

At 37 °C, fluorescent puncta moved towards the stereociliary tip. **f–h**, Immunogold electron microscopic images of the tip-link region of guinea-pig outer hair cells. **i**, Distances of gold particles from the lower insertion site of tip links, demonstrating overlap of labelling for PB811 and PB264. **j**, Diagram of CDH23 and PCDH15 showing the localization of the epitopes for the antibodies. **k**, Model of CDH23 and PCDH15 localization at tip links. Scale bars: **a–e**, 2.5 μm ; **f–h**, 50 nm.

formation between CDH23 and PCDH15 in the presence of 1 mM Ca^{2+} was inhibited by LaCl_3 but not by BaCl_2 (Fig. 3d).

To determine which EC domains promote interactions between CDH23 and PCDH15, we generated fusion proteins containing the N-terminal 11 EC domains or 3 EC domains of CDH23 (Fig. 3a). We did not generate constructs with one or two EC domains because studies with classical cadherins have shown that such constructs show low binding affinity²². The 11 or 3 EC domains of CDH23 interacted with PCDH15 (Fig. 3e). To verify interactions independently, we expressed the transmembrane version of PCDH15 in NIH3T3 cells and incubated non-permeabilized cells with CDH23–Fc (Fig. 4a). The cells were then permeabilized; PCDH15 was visualized with an antibody to the cytoplasmic domain, and CDH23–Fc with an Fc-specific antibody. CDH23–Fc bound cells expressing PCDH15 but not control cells. The pattern of CDH23–Fc binding overlapped with PCDH15 staining at the cell membrane. Intracellular vesicles containing PCDH15 did not bind CDH23–Fc because CDH23–Fc was added before permeabilization. Binding was also observed with the 11 N-terminal EC domains of CDH23, but not when CDH23–Fc was omitted (Fig. 4a).

To determine which PCDH15 domains mediate interactions with CDH23, we generated constructs consisting of the N-terminal 3 EC domains of PCDH15, but the proteins were unstable. Single amino acid changes in PCDH15 are linked to non-syndromic recessive deafness *DFNB23* (ref. 16). We engineered two *DFNB23*-associated mutations into PCDH15–Fc (Fig. 4b). A mutation in EC1, but not EC2, abolished interactions between PCDH15–Fc and CDH23–His (Fig. 4c). We conclude that EC1 of PCDH15 is critical for adhesion, and that *DFNB23* is caused in part by adhesion defects. Mutations such as those in EC2 may cause deafness by affecting the mechanical properties of tip links.

We next analysed a mix of CDH23–His and PCDH15–His by electron microscopy. Most molecules appeared as *cis*-homodimers,

but we consistently observed complexes with the proper dimensions for CDH23 homodimers interacting *in trans* with PCDH15 homodimers (Fig. 4d; see also Supplementary Fig. 6) to form a complex ~180 nm in length, which is in good agreement with the reported length of tip links (150–200 nm)^{9,11}. One end of the complex frequently branched, which we infer to be the end where CDH23 is located. These findings provide further evidence that CDH23 and PCDH15 interact at their N termini.

Collectively, our findings provide evidence that tip links are asymmetric adhesion complexes consisting of CDH23 and PCDH15 interacting at their N termini and forming the upper and lower part of tip links, respectively. CDH23 and PCDH15 interact at 0.1 mM Ca^{2+} , which is below the concentration necessary for stable interactions between classical cadherins²³. Critical Trp residues in the EC1 domain of classical cadherins^{19,24} are not conserved in CDH23, and only one Trp residue is present in EC1 of PCDH15, suggesting that the binding surface between CDH23 and PCDH15 is optimized for endolymph conditions. Our findings also suggest that the two ends of a tip link are functionally distinct. Myo1c, the adaptation motor for the mechanoelectrical transduction channel, co-immunoprecipitated with CDH23 (ref. 7) and is absent from stereocilia of CDH23-deficient *waltzer* mice²⁵, suggesting that Myo1c interacts with the upper tip-link end. The membrane at the lower tip-link end appears under tension and is pulled upwards⁹, raising the possibility that part of the gating spring may localize close to PCDH15. The localization of CDH23 and PCDH15 at kinociliary and transient lateral links in hair cells (Supplementary Fig. 2h, i) suggests that these linkages may also be formed by heterophilic interactions between the two cadherins. This molecular asymmetry might be important for translating extracellular signals into cytoskeletal changes that determine the polarity of the stereociliary bundles. Consistent with this model, bundle polarity is affected in mice with a mutation in *Pcdh15* (ref. 26).

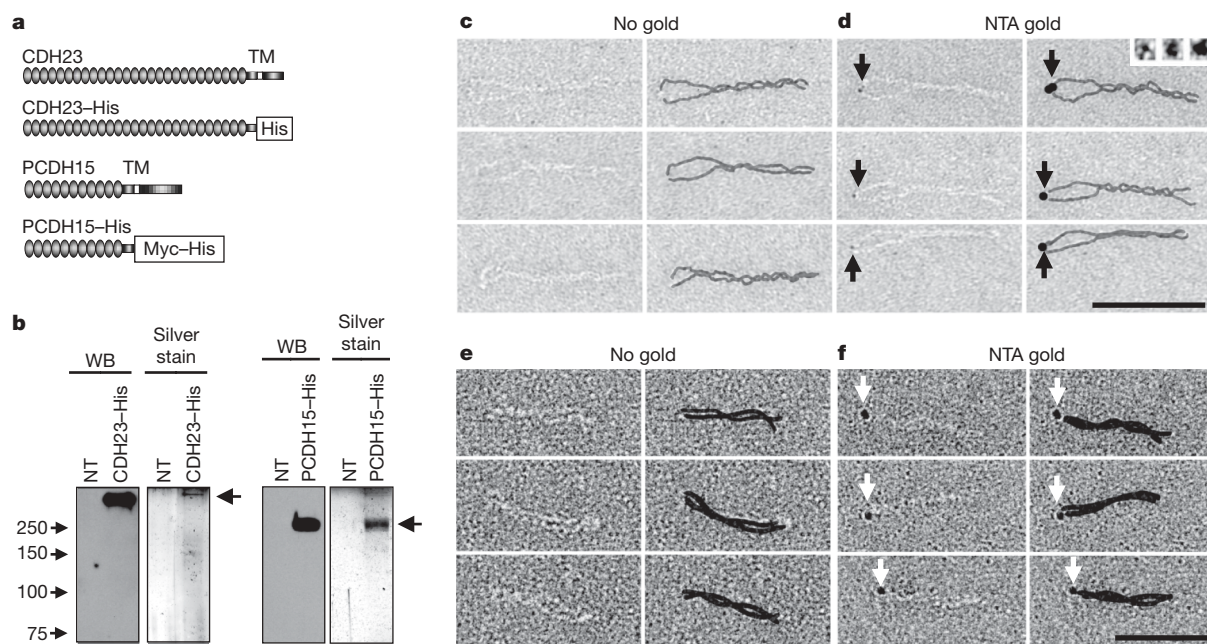


Figure 2 | The extracellular CDH23 and PCDH15 domains form parallel homodimers. **a**, Diagram of CDH23 and PCDH15 constructs. The extracellular domains were fused to a His tag or Myc–His tag. **b**, CDH23–His and PCDH15–His were affinity purified from the supernatant of transfected HEK293 cells. Full-length proteins were detected by western blot analysis and silver staining. No protein was visible in the supernatant of untransfected cells (NT). The molecular weight (kDa) is indicated on the left. **c–f**, Analysis of CDH23–His and PCDH15–His by negative staining

transmission electron microscopy. Strands are traced in the right panels for better visualization. **c**, CDH23–His formed dimers that frequently separated at one end. **d**, Ni-NTA nanogold beads (arrow) attached to the end of CDH23 homodimers that formed the branch/loop. Inset: sizes of a single and multiple nanogold beads (fivefold enlarged). In the right panels, gold beads are shown as black dots. **e**, PCDH15–His molecules formed dimers. **f**, Ni-NTA nanogold beads (arrow) attached to one end of the PCDH15–His homodimers. Scale bars: **c**, **d**, 100 nm; **e**, **f**, 50 nm.

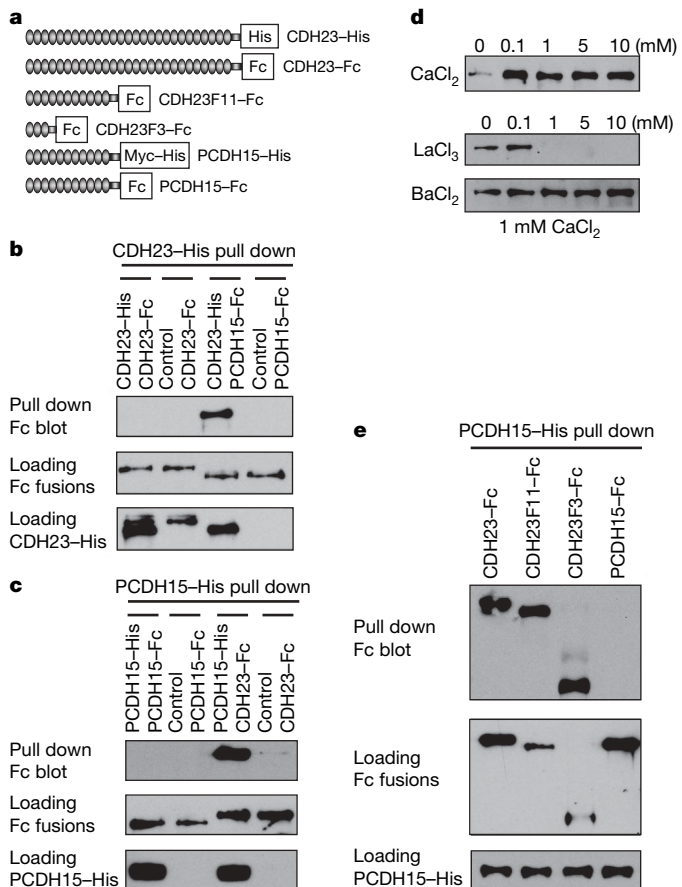


Figure 3 | CDH23 binds to PCDH15. **a**, Diagram of CDH23 and PCDH15 constructs. The proteins were purified from transfected HEK293 cells. **b**, CDH23-His was incubated with CDH23-Fc or PCDH15-Fc, protein complexes were purified and analysed by western blot analysis. In control lanes, CDH23-His was omitted. Top panel: CDH23-His binds PCDH15-Fc but not CDH23-Fc (detected with an Fc antibody). Middle and lower panels: loading controls for the Fc fusion proteins and CDH23 (detected with an antibody to Fc and CDH23, respectively). **c**, PCDH15-His was incubated with CDH23-Fc or PCDH15-Fc. In control lanes, PCDH15-His was omitted. The analysis of the protein complexes was carried out as described above. PCDH15-His bound to CDH23-Fc but not to PCDH15-Fc. **d**, Top panel: CDH23-His was incubated with PCDH15-Fc. Samples were washed in buffer containing the indicated amounts of CaCl_2 . Complex formation was observed in the presence of ≥ 0.1 mM CaCl_2 . Bottom panels: CDH23-His was incubated with PCDH15-Fc in the presence of 1 mM CaCl_2 and the indicated amounts of LaCl_3 or BaCl_2 . Complex formation was disrupted only by LaCl_3 . **e**, PCDH15-His was incubated with the indicated proteins; complexes were purified and analysed by western blotting. PCDH15-His bound to CDH23F11-Fc (N-terminal 11 EC domains) and CDH23F3-Fc (3 EC domains).

METHODS SUMMARY

Immunolocalization. Antibodies to CDH23 and PCDH15 were developed in rabbits and affinity purified. Antibodies PB264 and PB240 were raised against amino acids 123–135 (GDVNDNAPTFHNQ) and 1586–1597 (ATRPAPPDRERQ) of murine CDH23 (NM_023370), respectively. PB811 was raised against amino acids 80–96 (LSLKDNVDYWVLLDPVK) of PCDH15 (NP_075604). Immunofluorescence microscopy was performed with inner ear tissues from rats, mice and guinea-pigs as reported²⁷. For immunogold labelling, tissues from guinea-pigs (P90–P120) were fixed and labelled with primary antibody, washed, and incubated with 10 nm colloidal-gold-conjugated goat anti-rabbit IgG (BB International). Freeze-etching following immunogold labelling was carried out as described⁹.

Biochemistry. Expression constructs were generated by polymerase chain reaction (PCR) using described constructs^{7,26} and methods²⁸ and introduced in pCDNA3 (Invitrogen). The Fc domain was derived from hIgG1DNPARK (Genentech). HEK293 cells were transfected using FuGENE6 reagent

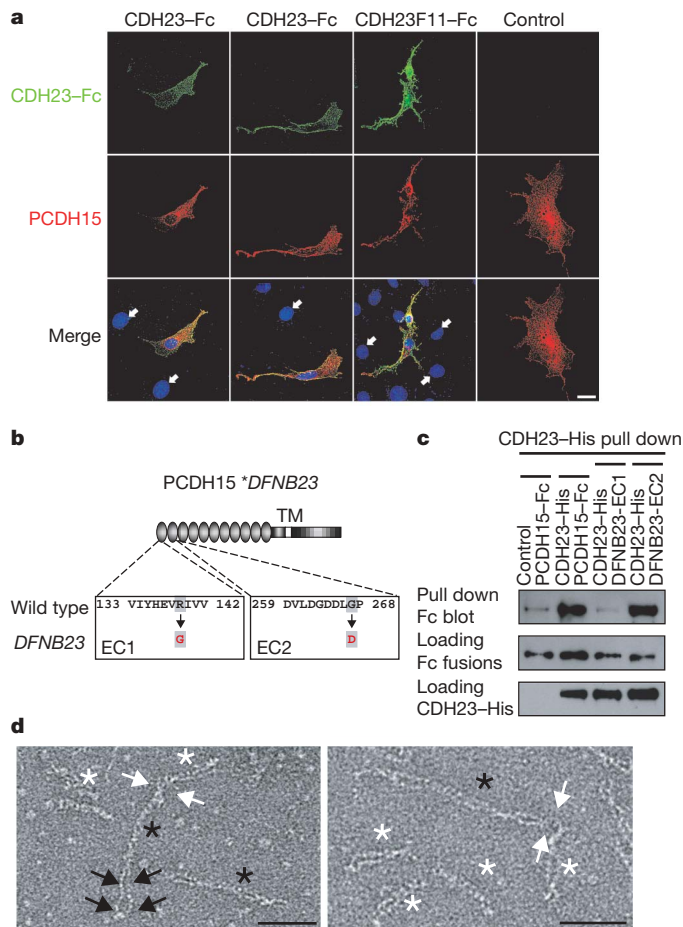


Figure 4 | Analysis of interactions between CDH23 and PCDH15. **a**, Full-length PCDH15 was expressed by transfection in NIH3T3 cells. Cells were incubated with recombinant CDH23-Fc and CDH23F11-Fc (indicated on the top of the panels). PCDH15 (red) and CDH23 (green) were visualized with antibodies coupled to fluorochromes. The merged panel also shows DAPI staining (blue). Only cells expressing PCDH15 bound CDH23-Fc and CDH23F11-Fc. No staining was observed when CDH23-Fc was omitted. **b**, PCDH15-Fc constructs with point mutations in EC1 and EC2 found in *DFNB23* patients¹⁶. **c**, CDH23-His was incubated with the indicated proteins, complexes were purified and analysed by western blot analysis. Top panel: CDH23-His binds to PCDH15-Fc and PCDH15(*DFNB23*-EC2), but not to PCDH15(*DFNB23*-EC1) (detected with an Fc antibody). Middle and lower panels: loading controls for the Fc fusions and CDH23-His (detected with an antibody to Fc and CDH23, respectively). **d**, CDH23-His and PCDH15-His were mixed and analysed by negative staining and electron microscopy. CDH23-His dimers, black asterisk; PCDH15-His bound to CDH23F11-Fc (N-terminal 11 EC domains) and CDH23F3-Fc (3 EC domains). Scale bars: **a**, 20 μm ; **d**, 50 nm.

(Roche), medium was collected and His-tagged proteins purified by affinity chromatography using Ni-NTA beads (Qiagen). Fc fusions were produced in a similar way. For structural studies, CDH23-His and PCDH15-His were applied to carbon grids, stained with 1% uranyl acetate and imaged using a Philips CM200 FEG TEM. For binding assays, His-tagged CDH23 was incubated with Fc-tagged PCDH15, protein complexes were purified with Ni-NTA beads and analysed by western blotting. CDH23 was detected with PB240, Fc-tagged proteins with horseradish peroxidase (HRP)-conjugated anti-human IgG (Amersham Biosciences), and PCDH15-His with HRP-conjugated anti-Myc antibody (Upstate Biotechnology). For immunocytochemistry, NIH3T3 cells were transfected with full-length PCDH15 (ref. 28), incubated with medium containing Fc-tagged CDH23, washed, fixed and stained with Alexa-488-conjugated anti-IgG antibody (Molecular Probes). Cells were permeabilized and re-stained with anti-PCDH15 cytodomain antibody²⁸, and with DAPI (Molecular Probes).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 April; accepted 10 July 2007.

1. Corey, D. Sensory transduction in the ear. *J. Cell Sci.* **116**, 1–3 (2003).
2. Gillespie, P. G. & Walker, R. G. Molecular basis of mechanosensory transduction. *Nature* **413**, 194–202 (2001).
3. Hudspeth, A. J. How hearing happens. *Neuron* **19**, 947–950 (1997).
4. Ahmed, Z. M. *et al.* The tip-link antigen, a protein associated with the transduction complex of sensory hair cells, is protocadherin-15. *J. Neurosci.* **26**, 7022–7034 (2006).
5. Lagziel, A. *et al.* Spatiotemporal pattern and isoforms of cadherin 23 in wild type and waltzer mice during inner ear hair cell development. *Dev. Biol.* **280**, 295–306 (2005).
6. Michel, V. *et al.* Cadherin 23 is a component of the transient lateral links in the developing hair bundles of cochlear sensory cells. *Dev. Biol.* **280**, 281–294 (2005).
7. Siemens, J. *et al.* Cadherin 23 is a component of the tip link in hair-cell stereocilia. *Nature* **428**, 950–955 (2004).
8. Sollner, C. *et al.* Mutations in cadherin 23 affect tip links in zebrafish sensory hair cells. *Nature* **428**, 955–959 (2004).
9. Kachar, B., Parakkal, M., Kurc, M., Zhao, Y. & Gillespie, P. G. High-resolution structure of hair-cell tip links. *Proc. Natl Acad. Sci. USA* **97**, 13336–13341 (2000).
10. Sotomayor, M., Corey, D. P. & Schulten, K. In search of the hair-cell gating spring elastic properties of ankyrin and cadherin repeats. *Structure* **13**, 669–682 (2005).
11. Tsuprun, V., Goodyear, R. J. & Richardson, G. P. The structure of tip links and kinociliary links in avian sensory hair bundles. *Biophys. J.* **87**, 4106–4112 (2004).
12. Ahmed, Z. M. *et al.* Mutations of the protocadherin gene *PCDH15* cause Usher syndrome type 1F. *Am. J. Hum. Genet.* **69**, 25–34 (2001).
13. Bolz, H. *et al.* Mutation of *CDH23*, encoding a new member of the cadherin gene family, causes Usher syndrome type 1D. *Nature Genet.* **27**, 108–112 (2001).
14. Bork, J. M. *et al.* Usher syndrome 1D and nonsyndromic autosomal recessive deafness DFNB12 are caused by allelic mutations of the novel cadherin-like gene *CDH23*. *Am. J. Hum. Genet.* **68**, 26–37 (2001).
15. Di Palma, F. *et al.* Mutations in *Cdh23*, encoding a new type of cadherin, cause stereocilia disorganization in waltzer, the mouse model for Usher syndrome type 1D. *Nature Genet.* **27**, 103–107 (2001).
16. Ahmed, Z. M. *et al.* *PCDH15* is expressed in the neurosensory epithelium of the eye and ear and mutant alleles are responsible for both USH1F and DFNB23. *Hum. Mol. Genet.* **12**, 3215–3223 (2003).
17. Assad, J. A., Shepherd, G. M. & Corey, D. P. Tip-link integrity and mechanical transduction in vertebrate hair cells. *Neuron* **7**, 985–994 (1991).
18. Halbleib, J. M. & Nelson, W. J. Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. *Genes Dev.* **20**, 3199–3214 (2006).
19. Boggon, T. J. *et al.* C-cadherin ectodomain structure and implications for cell adhesion mechanisms. *Science* **296**, 1308–1313 (2002).
20. Buchel, C., Morris, E., Orlova, E. & Barber, J. Localisation of the PsbH subunit in photosystem II: a new approach using labelling of His-tags with a Ni²⁺-NTA gold cluster and single particle analysis. *J. Mol. Biol.* **312**, 371–379 (2001).
21. Anniko, M. & Wroblewski, R. Ionic environment of cochlear hair cells. *Hear. Res.* **22**, 279–293 (1986).
22. Briehner, W. M., Yap, A. S. & Gumbiner, B. M. Lateral dimerization is required for the homophilic binding activity of C-cadherin. *J. Cell Biol.* **135**, 487–496 (1996).
23. Pokutta, S., Herrenknecht, K., Kemler, R. & Engel, J. Conformational changes of the recombinant extracellular domain of E-cadherin upon calcium binding. *Eur. J. Biochem.* **223**, 1019–1026 (1994).
24. Patel, S. D. *et al.* Type II cadherin ectodomain structures: implications for classical cadherin specificity. *Cell* **124**, 1255–1268 (2006).
25. Phillips, K. R., Tong, S., Goodyear, R., Richardson, G. P. & Cyr, J. L. Stereociliary myosin-1c receptors are sensitive to calcium chelation and absent from cadherin 23 mutant mice. *J. Neurosci.* **26**, 10777–10788 (2006).
26. Senften, M. *et al.* Physical and functional interaction between protocadherin 15 and myosin VIIa in mechanosensory hair cells. *J. Neurosci.* **26**, 2060–2071 (2006).
27. Schneider, M. E. *et al.* A new compartment at stereocilia tips defined by spatial and temporal patterns of myosin IIIa expression. *J. Neurosci.* **26**, 10243–10252 (2006).
28. Siemens, J. *et al.* The Usher syndrome proteins cadherin 23 and harmonin form a complex by means of PDZ-domain interactions. *Proc. Natl Acad. Sci. USA* **99**, 14946–14951 (2002).
29. Gillespie, P. G., Dumont, R. A. & Kachar, B. Have we found the tip link, transduction channel, and gating spring of the hair cell? *Curr. Opin. Neurobiol.* **15**, 389–396 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Sotomayor for discussions regarding cadherin ectodomain structure. We thank members of the Müller and Kachar laboratories for comments. This work was supported by the NIH (U.M., R.A.M., E.M.W.-K., and H.S., J.T., B.K. (intramural funding)). Negative staining TEM analysis was conducted at the National Resource of Automated Molecular Microscopy.

Author Contributions U.M. and B.K. supervised the project and contributed equally as co-senior authors. P.K. and H.S. are co-first authors. H.S. and J.T. characterized the antibodies and carried out the experiments in Fig. 1 and Supplementary Figs 1, 2 and 3. P.K. carried out the experiments in Figs 2–4 and Supplementary Figs 1, 4, 5 and 6. P.K. provided the proteins for negative staining TEM, which was carried out by E.M.W.-K.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to U.M. (umuell@scripps.edu) or B.K. (kacharb@nidcd.nih.gov).

LETTERS

Temporal precision in the neural code and the timescales of natural vision

Daniel A. Butts^{1,2}, Chong Weng³, Jianzhong Jin³, Chun-I Yeh^{3,4}, Nicholas A. Lesica¹, Jose-Manuel Alonso³ & Garrett B. Stanley¹

The timing of action potentials relative to sensory stimuli can be precise down to milliseconds in the visual system^{1–7}, even though the relevant timescales of natural vision are much slower. The existence of such precision contributes to a fundamental debate over the basis of the neural code and, specifically, what timescales are important for neural computation^{8–10}. Using recordings in the lateral geniculate nucleus, here we demonstrate that the relevant timescale of neuronal spike trains depends on the frequency content of the visual stimulus, and that ‘relative’, not absolute, precision is maintained both during spatially uniform white-noise visual stimuli and naturalistic movies. Using information-theoretic techniques, we demonstrate a clear role of relative precision, and show that the experimentally observed temporal structure in the neuronal response is necessary to represent accurately the more slowly changing visual world. By establishing a functional role of precision, we link visual neuron function on slow timescales to temporal structure in the response at faster timescales, and uncover a straightforward purpose of fine-timescale features of neuronal spike trains.

Figure 1a illustrates one of the many contexts in which millisecond precision has been observed in neuronal responses, showing the response of a geniculate neuron to repeated presentations of a spatially uniform white-noise visual stimulus (SUN). This remarkable precision at millisecond timescales has been observed in the retina^{2,7}, the lateral geniculate nucleus (LGN)^{5,6} and the visual cortex^{1,3,10} as well as in many other sensory systems such as the fly visual system^{4,9}, the electrosensory system of the weakly electric fish¹¹, and the mammalian somatosensory^{12,13} and auditory systems¹⁴. Although the presence of such fine temporal structure in the neuronal response would not be surprising if the sensory stimulus had similar temporal structure, its role is less clear in the mammalian visual system in which relevant visual stimuli are typically on much slower timescales. In particular, visual perception is ultimately limited by the relatively slow integration time of photoreceptors, which, for example, results in the appearance of continuous motion from the flickering images that constitute a movie. As a result, the much finer temporal structure in visual neuron responses has been proposed to be evidence for ‘temporal encoding’, which postulates that particular temporal patterns in the spike train carry additional information about the visual stimulus^{8,15}.

If millisecond temporal structure is important for the representation of visual information, one might expect it to be preserved in more natural stimulus conditions. Figure 1b shows the response of the same neuron in Fig. 1a responding to a ‘natural movie’ stimulus (see Methods)¹⁶. Although the response of the neuron still consists of discrete firing ‘events’, all relevant measures of event timing, ranging from the amount of trial-to-trial variability in timing of the first

spike (‘first-spike jitter’) to the overall event duration, are significantly slower during the natural movie versus during SUN (Supplementary Information 2). However, the increased response timescales are nearly matched by the increased timescales of the stimulus filtered by the receptive field of the neuron, because both the stimulus itself and the temporal filtering of the receptive field become slower in natural movies (Supplementary Information 1). By comparing the timescale of the ‘filtered stimulus’ τ_{FS} with that of the response τ_R during the SUN and the natural movie stimuli (Fig. 1c, d), we observe that ‘relative’ precision (defined as the ratio of the characteristic timescales of the filtered stimulus and neural response) is maintained at a level of 2–4 during both SUN and natural stimuli by all neurons in our study (Fig. 1e, f).

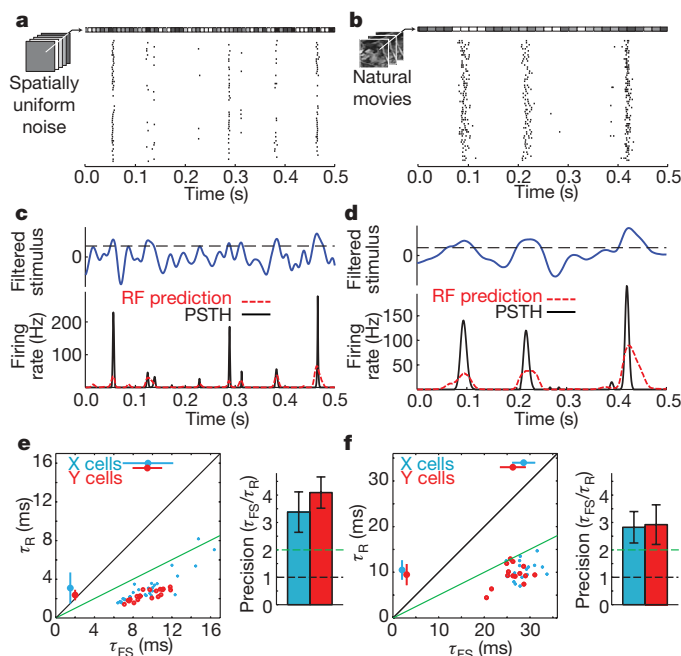


Figure 1 | The timescale of the neuronal response depends on the nature of the visual stimulus, defining relative precision. **a, b**, LGN X-cell responses (spike rasters) over 60 repeated trials in response to SUN (**a**) and natural movie stimuli (**b**). **c, d**, Corresponding time courses of the filtered stimulus (arbitrary units, with threshold shown as dashed line), PSTH (see Methods) and RF (receptive field) prediction. **e, f**, Left: comparison of the filtered stimulus and response timescales (τ_{FS} and τ_R , respectively) across the population of LGN neurons studied, for SUN (**e**, $n = 49$) and natural movies (**f**, $n = 32$). Right: the consistent ratio between these timescales (shown as mean \pm s.d.) defines the level of relative precision.

¹School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. ²The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Medical College of Cornell University, New York, New York 10021, USA. ³Department of Biological Sciences, State University of New York College of Optometry, New York, New York 10036, USA. ⁴Department of Psychology, University of Connecticut, Storrs, Connecticut 06269, USA.

Several information-based studies of retinal and geniculate spike trains in response to SUN have demonstrated that millisecond precision conveys additional information about the visual stimulus^{3–7}. However, we observe that—although still ‘precise’ in both stimulus contexts—the response timescale itself nearly quadruples from an average of 2.8 ms in the context of SUN to 10.1 ms in the natural movie. This suggests that any functional role of temporal precision cannot depend on fixed temporal features in the LGN response, and raises the question of whether the extra information that is encoded at millisecond timescales disappears altogether in more natural contexts. To address this question, we must first understand what this millisecond-timescale information represents about the visual stimulus. We investigated this using a ‘realistic model’ that reproduces the observed temporal features and trial-to-trial variability of LGN spike trains in the context of SUN (see Supplementary Information 3)¹⁷. Both the timescale of this model’s filtered stimulus (7.1 ms) and its response (1.6 ms) are in the range of experimental observations (Fig. 1e), giving a relative precision of 4.5; its spike train also carries a typical amount of information⁶.

Notably, the relative precision of the model can be varied systematically (Fig. 2a) from infinite precision (blue) to a relative precision of 4.5 (green; realistic model) and to a relative precision of 1 (red; matching the response timescale to the filtered stimulus timescale), without changing how the stimulus itself is processed by the receptive field. We evaluate the effects of precision on the information content

of the spike train using the single-spike information rate I_{SS} (ref. 18), which is a first-order approximation to the full mutual information^{6,19} and reflects the more general dependence of mutual information measures on response timescales. By systematically varying the precision, this analysis reveals what at first seems to be a paradoxical effect of precision: information about the stimulus grows without bound with increasing precision (Fig. 2b, note the log-scaling of the horizontal axis), even though no information can be created through spike generation that does not already exist at the level of the filtered stimulus²⁰.

To understand the nature of this literally boundless amount of information that becomes available with increasing precision, consider that, for a temporally varying stimulus, there is information to be gained in distinguishing between a stimulus $s(t)$ and its temporally shifted ‘twin’ $s(t + \Delta t)$ for any time shift Δt , regardless of the scale of the stimulus features (Fig. 2c). Despite this conceptual distinction between stimuli, as Δt becomes vanishingly small, $s(t) \approx s(t + \Delta t)$.

To formalize this intuition, it is necessary to apply more traditional measures of information between stimulus and response, based on linear stimulus reconstruction. The quality of the reconstruction is barely affected by the level of variability of the realistic model (Fig. 2d, middle), whereas increasing the temporal jitter such that the relative precision is decreased to 1 significantly degrades the reconstruction (Fig. 2d, bottom). This can be quantified by the reconstruction information I_{REC} (refs 8, 21 and 22), which explicitly measures how much information about the stimulus can be recovered from the observed neuronal response using linear reconstruction. Unlike the single-spike information, the reconstruction information saturates at high precision (Fig. 2e), such that the response precision of the realistic model (green dashed line) is sufficiently high to capture the available I_{REC} .

What is surprising is that, although the stimulus reconstruction is based on the receptive field function at the timescales of the filtered stimulus (Fig. 2e, red dashed line), we see that the neuron’s response must be significantly more precise to recover the full reconstruction information. Conversely, nearly half of the available information is lost if the response timescale is only dictated by the filtered stimulus. Thus, these results recapitulate the dichotomy between the timescales of receptive field function and the precision of the response.

As Fig. 2c implies, the frequency content of the stimulus determines the temporal scale at which the response must be specified to reconstruct the stimulus faithfully. The reconstruction information can be explicitly decomposed as a function of frequency using the coherence $\gamma^2(f)$ between the stimulus and the reconstructed stimulus. The coherence for the precise model neuron is compared to the coherence when the model’s precision is degraded by adding temporal jitter to its spike train (Fig. 3a). Decreased precision has the largest effect on the highest stimulus frequencies represented, and a simple relationship emerges when considering the relationship between the coherence of the precise model spike trains and the coherence of the jittered spike trains (Fig. 3b). As derived in Supplementary Information 4, the frequency scale of the degradation of the reconstruction (denoted here as the ‘attenuation frequency’, f_A) is directly related to the amount of temporal jitter in the neural response. This relationship provides a general guideline for understanding which filtered stimulus frequencies can be reconstructed for a certain amount of response jitter. For example, an LGN neuron recorded in the context of the natural movie has a filtered stimulus with frequency content up to 20 Hz (Fig. 3c and Supplementary Information 4). This frequency maps to a jitter of 6.2 ms (12.4-ms timescale), which nearly matches the observed timescale of the neuron’s response (Fig. 3d). Furthermore, the experimentally measured I_{REC} degrades for larger magnitudes of jitter, suggesting that the precision observed during the natural movie is just enough to avoid disrupting the information available to linear stimulus reconstruction.

Owing to the spatiotemporal nature of LGN receptive fields, the requirement for temporal precision also applies to the ability to

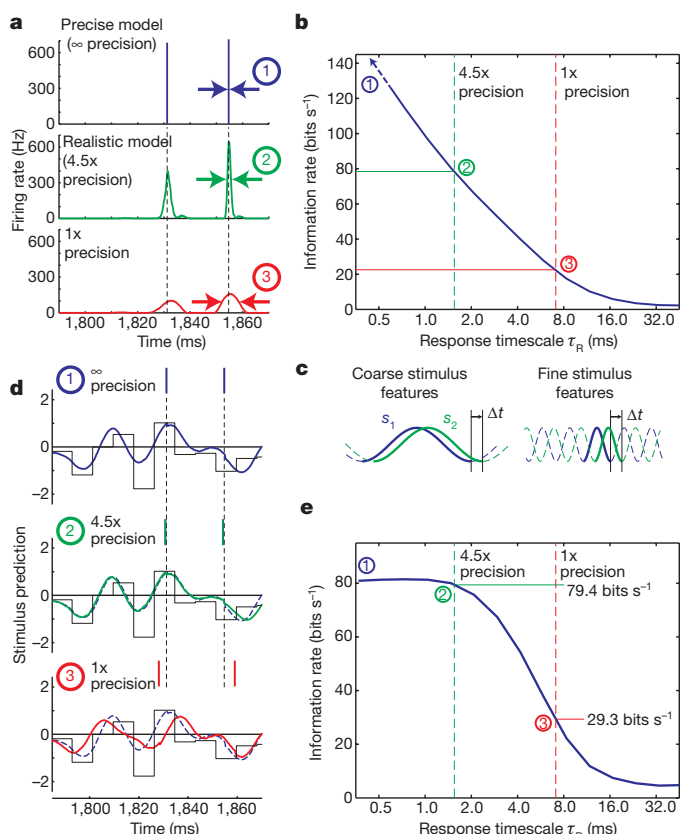


Figure 2 | Precision is necessary to convey information about visual stimuli. **a**, Resulting PSTH as the temporal precision of the LGN spike train is systematically varied. **b**, Single-spike information I_{SS} increases without bound with increasing precision. **c**, Jittering a reconstructed stimulus by Δt has a greater effect for fine (right) than for coarse (left) stimulus features. **d**, Jitter affects the reconstruction quality at a relative precision of 1 (bottom), but has negligible influence at a relative precision of 4.5 (middle). The stimulus is scaled such that its variance over time is unity, establishing the scale of the y axis. **e**, I_{REC} reveals that response timescales much finer than that of the stimulus (red) are necessary to maximize the neuron’s linear representation of the stimulus.

represent spatial aspects of the stimulus. We consider a simple spatiotemporal reconstruction using an array of model neurons, each representing one pixel of the natural movie stimulus considered in this paper¹⁶. Figure 3e shows single frames of the reconstructed movie from this simulation using temporally precise neurons (left), neurons with a relative precision of 2.3 (6.2-ms jitter, middle) and neurons with a relative precision of 1 (14.5-ms jitter, right). Although this stimulus only updates at 30 Hz (corresponding to 33-ms stimulus frames), the degradation of the quality in the image from left to right demonstrates that spatial information about the stimulus is represented at timescales finer than that of the stimulus. Furthermore, at the spatial scale of individual receptive fields (Fig. 3f), a simple 'edge detector' demonstrates that boundaries of objects present in the image become disrupted as precision decreases to 1. This has clear implications for the necessary temporal sensitivity of cortical neurons that receive direct input from the LGN, which are thought to encode edges. To have access to the full amount of information from their LGN inputs, they must be sensitive to these inputs at the same level of precision, which in the natural movie is on the order of 10 ms.

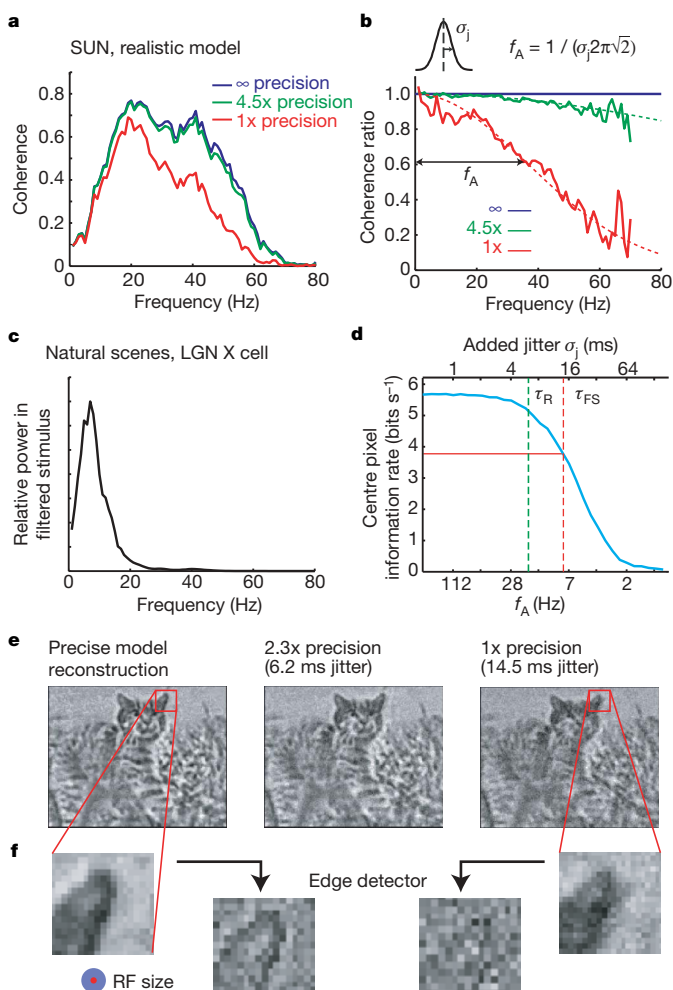


Figure 3 | Precision is necessary to represent relevant stimulus frequencies. **a**, Coherence between SUN and its reconstruction in models with varying amounts of jitter. **b**, Ratio of coherence between jittered and precise models, defining the attenuation frequency f_A . **c**, Power in the natural movie filtered by a measured X-cell receptive field. **d**, Reconstruction information for the X cell at its centre pixel as a function of added jitter (green, neuron's measured jitter, relative precision of 2.3; red, degraded, relative precision of 1). **e**, Frame of the reconstructed stimulus generated from an array of neurons with different precisions. **f**, Spatial fidelity is compromised by a lack of precision.

Thus, we have demonstrated that the 'classical' function of visual neurons—signalling stimuli in their receptive field—actually necessitates the neural precision observed experimentally. Whereas it has long been known that the timescales of visual neuron responses can be modulated by the nature of the visual stimulus, the analysis described here places this finding in the context of their functional role: representation of the natural visual scene. This provides a link between the timescales of natural vision (with frequency content largely <20 Hz owing to the filtering of LGN receptive fields, see Supplementary Information 4) and response timescales on the order of 10 ms, which are increasingly being observed in the cortex (for example, cortical dendritic integration²³, cortical synaptic plasticity²⁴, spike-train-based discrimination^{25,26} and gamma-band modulation of cortical fields²⁷). In this light, LGN inputs to the cortex may be at a suitable timescale for cortical computation, and understanding the temporal structure of these inputs at the appropriate level of precision may lead to a new perspective on cortical processing.

The observation that the timescales of neural responses are not absolute indicates that downstream neurons cannot depend on fixed temporal features of the response to encode the stimulus. Instead, we demonstrate that the information in the spike train (at least at the level of single spikes) is accessible through simple linear decoding, which relies on relative precision rather than fixed temporal patterning. Although this study does not preclude the existence of more complicated explanations for visual neuron precision that extend beyond receptive-field function (that is, 'temporal encoding'¹⁵) or across populations^{22,28,29}, it demonstrates that precision is not by itself evidence for such ideas. In this sense, our results establish stronger criteria that 'temporal encoding' must satisfy to demonstrate a functional role above and beyond straightforward receptive-field-based processing.

Relative precision may be a general feature of sensory neuron communication, in which an analogue input (the sensory stimulus) is encoded by what is essentially a digital signal (the neuron's spike train). In this context, temporal precision of neuronal responses is conceptually similar to the problem of digital sampling, in which encoding frequencies must be at least double that of the analogue signal information because of the Nyquist limit^{8,30}. From this perspective, the mechanisms that generate neuronal precision (Supplementary Information 3), which seem to make the encoding of visual information more complicated, may actually serve to provide easier means for downstream neurons to decode this information.

METHODS SUMMARY

Data were recorded extracellularly from layer A of the LGN of anaesthetized paralysed cats. Individual units were classified as X or Y according to their responses to counterphase sinusoidal gratings.

The SUN stimuli were 120 Hz gaussian white noise with an r.m.s. (root-mean-squared) contrast of 0.55. The naturalistic movie was generated by a camera mounted to the head of a cat roaming in a forest, provided by the König laboratory (Institute of Neuroinformatics, ETH/UNI Zürich, Switzerland)¹⁶. It was modified so that each frame had the same mean luminance and an r.m.s. contrast of 0.4, and was updated at 60 Hz. All stimuli were presented on a cathode-ray tube display with a 120 Hz monitor refresh rate; we verified that the monitor refresh did not affect our results (Supplementary Information 5).

The receptive field predictions were based on 'linear-non-linear' (LN) models, estimated using standard techniques (Supplementary Information 1). The filtered stimulus is the linear convolution of the stimulus with the receptive field. Filtered stimulus and response peri-stimulus time histogram (PSTH) timescales (τ_{FS} and τ_R) were derived from their autocorrelation functions (Supplementary Information 1).

The parameters of the realistic model LGN neuron were provided by M. Meister¹⁷. Precise model responses were generated from the realistic model by identifying individual spikes across trials and shifting each spike to its average time over all trials in which it was present. The amount of each shift was then scaled to change the level of precision (Fig. 2). The spatiotemporal simulation (Fig. 3e, f) consisted of realistic model neurons modified to have measured spatiotemporal receptive fields (Supplementary Information 1) and assembled into two 160 × 120 overlapping arrays (ON and OFF). Edge detection (Fig. 3f)

was performed by measuring the difference between each pixel and the average of the surrounding pixels.

The single-spike information rate¹⁸ was calculated directly from the PSTH, and the reconstruction information I_{REC} was calculated from the coherence $\gamma^2(f)$ between the stimulus and the reconstructed stimulus: $I_{\text{REC}} = -\int df \log_2[1 - \gamma^2(f)]$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 February; accepted 16 July 2007.

- Bair, W. & Koch, C. Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Comput.* **8**, 1185–1202 (1996).
- Berry, M. J. II & Meister, M. Refractoriness and neural precision. *J. Neurosci.* **18**, 2200–2211 (1998).
- Buracas, G. T., Zador, A. M., DeWeese, M. R. & Albright, T. D. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron* **20**, 959–969 (1998).
- Lewen, G. D., Bialek, W. & de Ruyter van Steveninck, R. R. Neural coding of naturalistic motion stimuli. *Network* **12**, 317–329 (2001).
- Liu, R. C., Tzovev, S., Rebrik, S. & Miller, K. D. Variability and information in a neural code of the cat lateral geniculate nucleus. *J. Neurophysiol.* **86**, 2789–2806 (2001).
- Reinagel, P. & Reid, R. C. Temporal coding of visual information in the thalamus. *J. Neurosci.* **20**, 5392–5400 (2000).
- Uzzell, V. J. & Chichilnisky, E. J. Precision of spike trains in primate retinal ganglion cells. *J. Neurophysiol.* **92**, 780–789 (2004).
- Borst, A. & Theunissen, F. E. Information theory and neural coding. *Nature Neurosci.* **2**, 947–957 (1999).
- de Ruyter van Steveninck, R. R., Lewen, G. D., Strong, S. P., Koberle, R. & Bialek, W. Reproducibility and variability in neural spike trains. *Science* **275**, 1805–1808 (1997).
- Mainen, Z. F. & Sejnowski, T. J. Reliability of spike timing in neocortical neurons. *Science* **268**, 1503–1506 (1995).
- Gabbiani, F., Metzner, W., Wessel, R. & Koch, C. From stimulus encoding to feature extraction in weakly electric fish. *Nature* **384**, 564–567 (1996).
- Bolouri, A. R. & Stanley, G. B. The dynamics of spatiotemporal response integration in the somatosensory cortex of the vibrissa system. *J. Neurosci.* **26**, 3767–3782 (2006).
- Phillips, J. R., Johnson, K. O. & Hsiao, S. S. Spatial pattern representation and transformation in monkey somatosensory cortex. *Proc. Natl Acad. Sci. USA* **85**, 1317–1321 (1988).
- Wehr, M. & Zador, A. M. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426**, 442–446 (2003).
- Theunissen, F. & Miller, J. P. Temporal encoding in nervous systems: a rigorous definition. *J. Comput. Neurosci.* **2**, 149–162 (1995).
- Kayser, C., Salazar, R. F. & Konig, P. Responses to natural scenes in cat V1. *J. Neurophysiol.* **90**, 1910–1920 (2003).
- Keat, J., Reinagel, P., Reid, R. C. & Meister, M. Predicting every spike: a model for the responses of visual neurons. *Neuron* **30**, 803–817 (2001).
- Brenner, N., Strong, S. P., Koberle, R., Bialek, W. & de Ruyter van Steveninck, R. R. Synergy in a neural code. *Neural Comput.* **12**, 1531–1552 (2000).
- Pola, G., Thiele, A., Hoffmann, K. P. & Panzeri, S. An exact method to quantify the information transmitted by different mechanisms of correlational coding. *Network* **14**, 35–60 (2003).
- de Ruyter van Steveninck, R. & Laughlin, S. B. The rate of information transfer at graded-potential synapses. *Nature* **379**, 642–645 (1996).
- Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R. & Warland, D. Reading a neural code. *Science* **252**, 1854–1857 (1991).
- Dan, Y., Alonso, J. M., Usrey, W. M. & Reid, R. C. Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nature Neurosci.* **1**, 501–507 (1998).
- Yuste, R. & Denk, W. Dendritic spines as basic functional units of neuronal integration. *Nature* **375**, 682–684 (1995).
- Fu, Y. X. *et al.* Temporal specificity in the cortical plasticity of visual space representation. *Science* **296**, 1999–2003 (2002).
- Narayan, R., Grana, G. & Sen, K. Distinct time scales in cortical discrimination of natural sounds in songbirds. *J. Neurophysiol.* **96**, 252–258 (2006).
- Chichilnisky, E. J. & Kalmar, R. S. Temporal resolution of ensemble visual motion signals in primate retina. *J. Neurosci.* **23**, 6681–6689 (2003).
- Womelsdorf, T., Fries, P., Mitra, P. P. & Desimone, R. Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature* **439**, 733–736 (2006).
- Ahissar, E. & Arieli, A. Figuring space by time. *Neuron* **32**, 185–201 (2001).
- VanRullen, R., Guyonneau, R. & Thorpe, S. J. Spike times make sense. *Trends Neurosci.* **28**, 1–4 (2005).
- Lazar, A. A. Perfect recovery and sensitivity analysis of time encoded bandlimited signals. *IEEE Trans. Circ. Syst.* **51**, 2060–2073 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a Charles King Trust Postdoctoral Fellowship (Bank of America, Co-Trustee, Boston; D.A.B.), by the NGIA (D.A.B., N.A.L., G.B.S.), by the NIH and by the SUNY Research Foundation (C.W., J.J., C.-I.Y., J.-M.A.). We thank M. Goldman, M. Meister, G. Desbordes and A. Bolouri for comments on the manuscript, C. Kayser for providing the natural-scene movies, and P. Wolfe for discussions regarding sampling issues.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.A.B. (dab2024@med.cornell.edu).

METHODS

Experimental procedures. Surgical and experimental procedures were performed in accordance with United States Department of Agriculture guidelines and were approved by the Institutional Animal Care and Use Committee at the State University of New York, State College of Optometry. As described in detail in ref. 31, cats were initially anaesthetized with ketamine (10 mg kg^{-1} intramuscular) followed by thiopental sodium (20 mg kg^{-1} intravenous during surgery and at a continuous rate of $1\text{--}2 \text{ mg kg}^{-1} \text{ h}^{-1}$ intravenous during recording; supplemented as needed). A craniotomy and duratomy were performed to introduce recording electrodes into the LGN (anterior, 5.5; lateral, 10.5). Animals were paralysed with atracurium besylate ($0.6\text{--}1 \text{ mg kg}^{-1} \text{ h}^{-1}$ intravenous) to minimize eye movements, and were artificially ventilated. LGN responses were recorded extracellularly within layer A. Recorded voltage signals were conventionally amplified, filtered and passed to a computer running the RASPUTIN software package (Plexon). For each cell, spike waveforms were identified initially during the experiment and were verified carefully off-line by spike-sorting analysis. Cells were classified as X or Y according to their responses to counter-phase sinusoidal gratings. Cells were eliminated from this study if they did not have at least 2 Hz mean firing rates in response to all stimulus conditions, or if the maximum amplitude of their spike-triggered average in spatiotemporal white noise was not at least five times greater than the amplitude outside of the receptive field area.

Visual stimuli. SUN stimuli consisted of spatially uniform luminances randomly selected from a gaussian distribution with zero mean (corresponding to the midway point of the full range of monitor luminance) and a root-mean-squared (r.m.s.) contrast of 0.55, presented at 120 Hz. The naturalistic movie sequence was recorded using a removable lightweight CCD camera mounted to the head of a freely roaming cat in natural environments such as grassland and forest¹⁶. A 48×48 windowed area was processed to play at 60 Hz and to have a constant mean and standard deviation of luminance for each frame (contrast held at 0.4 of maximum)³². All stimuli were displayed on a cathode-ray tube display at a resolution of 0.2° per pixel with a monitor refresh rate of 120 Hz. Care was taken to ensure that the potential effects of phase-locking to the monitor refresh did not affect the results (Supplementary Information 5).

Receptive-field-based predictions and the filtered stimulus. The receptive field predictions (Fig. 1c, d) were based on linear–non-linear (LN) models that were estimated using standard techniques (see Supplementary Information 1). Temporal-only receptive fields were estimated directly from responses to SUN, whereas we used 60 Hz spatiotemporal binary noise to map the spatiotemporal receptive field. The filtered stimulus (for example, Fig. 1c, d) is defined as the linear convolution of the stimulus with the receptive field, and represents how similar the stimulus is to the receptive field at a particular time. It is scaled to have a standard deviation of unity for a given stimulus class, and is mapped to a firing

rate through a ‘static non-linearity’ also using standard techniques (Supplementary Information 1).

Timescale of filtered stimulus and PSTH. The PSTH was estimated from 60 or 120 repeated trials, and was subsequently used to estimate response timescales and single-spike information (Supplementary Information 1).

The ‘realistic’ and ‘precise’ models. The parameters of the realistic model used in this paper were provided by M. Meister, originating from a fit to a representative cat LGN cell (neuron number 6 in ref. 17) recorded under similar experimental conditions to those used here. The precise model responses were generated from the realistic model spike times over 1,000 repeated trials by identifying individual spikes across trials and shifting each spike to agree with its average time over all trials in which it appeared. The amount of each shift of each spike from the precise model was then scaled to generate models with any amount of precision (Fig. 2).

Linear reconstruction. The optimal linear reconstruction filter $h(\tau)$ corresponds to the temporal filter that minimizes the mean-squared difference between the stimulus $s(t)$ and the reconstructed stimulus $s^*(t)$ (refs 8 and 21). The reconstruction filters were estimated numerically for a given stimulus and set of experimentally observed spike times.

Information measurements. The single spike information rate¹⁸ is given by:

$$I_{\text{SS}}[S, \lambda] = \lambda_0 \int dt \left[\frac{\lambda(t)}{\lambda_0} \right] \log_2 \left[\frac{\lambda(t)}{\lambda_0} \right] \quad (1)$$

In equation (1), $\lambda(t)$ is the instantaneous firing rate characterized by the PSTH, and λ_0 is the mean firing rate. The reconstruction information I_{REC} is calculated from the coherence spectra $\gamma^2(f)$ between stimulus $s(t)$ and reconstructed stimulus $s^*(t)$: $I_{\text{REC}} = -\int df \log_2[1 - \gamma^2(f)]$. In the case of the natural movie reconstruction, only the reconstruction information of the centre pixel of the receptive field is reported.

LGN model neurons and natural movie simulation. We used a 160×120 array of precise neurons with the same parameters as the precise model of Fig. 2, except we replaced the purely temporal receptive field with a measured spatiotemporal receptive field. The spatial components were spatially shifted for each neuron to tile a 160×120 spatiotemporal stimulus derived from the same natural movie. An ON centre cell and an OFF centre cell were simulated at each position, and each pixel was independently reconstructed into a 160×120 image evolving over time using the resulting spike trains.

31. Weng, C., Yeh, C. I., Stoelzel, C. R. & Alonso, J. M. Receptive field size and response latency are correlated within the cat visual thalamus. *J. Neurophysiol.* **93**, 3537–3547 (2005).
32. Lesica, N. A. *et al.* Dynamic encoding of natural luminance sequences by LGN bursts. *PLoS Biol.* **4**, e209 (2006).

LETTERS

Polo inhibits progenitor self-renewal and regulates Numb asymmetry by phosphorylating Pon

Hongyan Wang^{1*}, Yingshi Ouyang^{2*}, W. Gregory Somers¹, William Chia¹ & Bingwei Lu²

Self-renewal and differentiation are cardinal features of stem cells. Asymmetric cell division provides one fundamental mechanism by which stem cell self-renewal and differentiation are balanced^{1,2}. A failure of this balance could lead to diseases such as cancer^{3–6}. During asymmetric division of stem cells, factors controlling their self-renewal and differentiation are unequally segregated between daughter cells. Numb is one such factor that is segregated to the differentiating daughter cell during the stem-cell-like neuroblast divisions in *Drosophila melanogaster*⁷, where it inhibits self-renewal^{8,9}. The localization and function of Numb is cell-cycle-dependent^{7,10–12}. Here we show that Polo (ref. 13), a key cell cycle regulator, the mammalian counterparts of which have been implicated as oncogenes as well as tumour suppressors^{14,15}, acts as a tumour suppressor in the larval brain. Supernumerary neuroblasts are produced at the expense of neurons in *polo* mutants. Polo directly phosphorylates Partner of Numb (Pon, ref. 16), an adaptor protein for Numb, and this phosphorylation event is important for Pon to localize Numb. In *polo* mutants, the asymmetric localization of Pon, Numb and atypical protein kinase C are disrupted, whereas other polarity markers are largely unaffected. Overexpression of Numb suppresses neuroblast overproliferation caused by *polo* mutations, suggesting that Numb has a major role in mediating this effect of Polo. Our results reveal a biochemical link between the cell cycle and the asymmetric protein localization machinery, and indicate that Polo can inhibit progenitor self-renewal by regulating the localization and function of Numb.

Asymmetric localization of Numb depends on its adaptor protein Pon. The Pon localization domain (Pon-LD) is located at the carboxy terminus of the protein¹⁷. The Ser 611 (S611) residue in this domain matches the consensus phosphorylation site for Polo (Fig. 1a), a principal orchestrator of cell cycle events. Because the localization of Pon is cell-cycle-dependent, we tested whether Polo can directly phosphorylate Pon. Pon-LD, but not Pon(S611A)-LD, in which S611 was mutated to Ala, was readily phosphorylated by mammalian Polo-like kinase 1 (Plk1) *in vitro* (Fig. 1b), demonstrating that Pon S611 is a Polo phosphorylation site.

To test whether Pon S611 is normally phosphorylated *in vivo*, we generated an antibody against S611-phosphorylated (p-S611) Pon. The specificity of this antibody was shown by its ability to recognize a glutathione S-transferase–Pon-LD fusion protein (GST–Pon-LD) only after the fusion protein was pre-phosphorylated by Plk1 (Fig. 1c). It did not recognize GST–Pon(S611A)-LD in the same assay. Next, larval brain extracts prepared from wild type as well as heterozygotes (*polo*⁹(+/-) and *polo*¹⁰(+/-)), and homozygotes (*polo*⁹(-/-) and *polo*¹⁰(-/-)) of two different *polo* loss-of-function alleles were analysed by western blotting using this p-S611-specific antibody. p-S611-positive Pon was clearly detected in wild-type

animals and in heterozygotes, but was barely detectable in homozygous mutant animals (Fig. 1d), demonstrating that Pon is phosphorylated at S611 *in vivo* in a Polo-dependent fashion.

We also used immunohistochemistry to verify S611 phosphorylation *in vivo* and to visualize phospho-Pon localization. p-S611-positive endogenous Pon was detected in wild-type larval brains

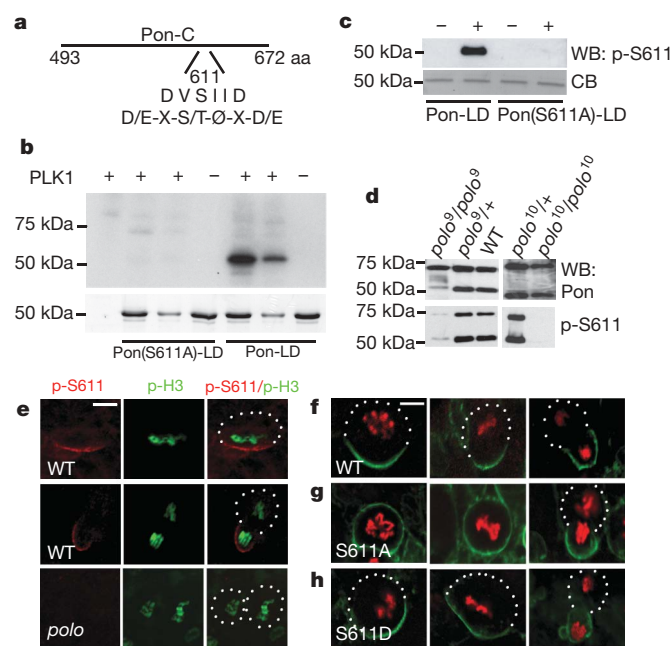


Figure 1 | Polo phosphorylates Pon and this phosphorylation event regulates the asymmetric localization of Pon. **a**, A diagram of the putative Polo phosphorylation site (S611) in Pon-LD, which matches the consensus Polo phosphorylation site (D/E-X-S/T-Ø-X-D/E, where Ø is a hydrophobic amino acid and X is any amino acid). **b**, *In vitro* kinase assays. Upper panel, autoradiography; lower panel, Coomassie blue (CB) staining of the same gel. **c**, Western blot (WB) analysis testing the specificity of the p-S611 antibody. **d**, Western blot analysis showing loss of p-S611-positive Pon in *polo* mutants. Note that, in addition to the 72 kDa full-length Pon protein, both the general Pon antibody and the p-S611 antibody recognized a ~50 kDa band—a putative proteolytic product. The level of the 50 kDa band seems to be dependent on Polo activity. **e**, Staining of endogenous Pon in wild-type (WT, upper and middle panels) and *polo* mutant (lower panels) cells using the p-S611 antibody, p-H3, p-histone H3. **f–h**, Localization of GFP–Pon-LD (WT) (**f**), GFP–Pon(S611A)-LD (**g**) or GFP–Pon(S611D)-LD (**h**) in embryos at stages 9–12. Neuroblasts at different cell cycle stages are shown. GFP–Pon-LD proteins were detected by immunostaining with anti-myc antibody (green). Anti-p-histone H3 staining (red) was used to indicate mitotic stages. Scale bars, 5 µm.

¹Temasek Life Sciences Laboratory and Department of Biological Sciences, National University of Singapore, Singapore 117604. ²Department of Pathology, Stanford University School of Medicine, Geriatric Research, Education and Clinical Center (GRECC)/VA Palo Alto Health Care System (VAPAHCS), Palo Alto, California 94304, USA.

*These authors contributed equally to this work.

as a crescent in metaphase neuroblasts, and was segregated to the ganglion mother cell (GMC, the daughter committed to the differentiation pathway) after division (Fig. 1e, upper and middle panels; $n = 58$, 100%). In the *polo*⁹ mutant, however, p-S611-positive Pon was undetectable (Fig. 1e, lower panels; $n = 43$, 100%). The p-S611 antibody also reacted with Pon-LD, but not with Pon(S611A)-LD, in transgenic larval brain (Supplementary Fig. 1).

To test for a functional role of S611 phosphorylation, we mutated S611 to a non-phosphorylatable Ala residue (S611A) or to a phospho-mimetic Asp residue (S611D). Wild-type and phospho-mutant Pon-LD were fused to green fluorescent protein (GFP) and expressed in embryonic neuroblasts. Both GFP-Pon-LD (Fig. 1f; $n = 35$, 100%) and GFP-Pon(S611D)-LD (Fig. 1h; $n = 35$, 100%) showed the expected basal localization. In contrast, the localization of GFP-Pon(S611A)-LD was defective. At prometaphase and metaphase, it showed either uniformly cortical (80%) or basally enriched but apically detectable cortical (20%) distribution (Fig. 1g, left and middle panels; $n = 56$). At anaphase and telophase, however, it formed basal crescents in most neuroblasts (Fig. 1g, right panel; $n = 42$, 93%). This 'telophase rescue' seemed to be partially mediated by endogenous Pon, because less rescue was observed in *pon* mutant neuroblasts, with 58% ($n = 17$) neuroblasts mis-segregating GFP-Pon(S611A)-LD at late anaphase/telophase (Supplementary Fig. 2). It is unlikely that the S611A mutation affects Pon localization by altering the charge or global conformation and folding of the protein, because mutation of an adjacent Ser residue (S616) or triple mutations at three potential atypical protein kinase C (aPKC) phosphorylation sites (S644A/S648A/S652A) had no effect on Pon-LD localization (Supplementary Fig. 3), suggesting that S611 represents a unique regulatory point in Pon localization.

To assess whether Polo has a role in neuroblast self-renewal and/or asymmetric division, we quantified central brain neuroblast numbers in two strong hypomorphic alleles, *polo*⁹ and *polo*¹⁰, using Deadpan (Dpn) and Miranda (Mira) as neuroblast markers. Wild-type larval central brains contained 37.3 ± 5.5 neuroblasts 24 h after larval hatching (ALH) and 90.7 ± 7.0 neuroblasts 96 h ALH (Fig. 2a). *polo*⁹ larval central brains had 36.5 ± 4.5 neuroblasts 24 h ALH. However, the number increased dramatically to 253.6 ± 44.5 neuroblasts 96 h ALH (Fig. 2a). Consistent with this increase in neuroblast number (Fig. 2a, and lower panels of b–d), the numbers of BrdU-labelled (lower panel of Fig. 2g), CycE-positive (lower panel of Fig. 2h) or phospho-histone-H3-positive proliferating cells (lower panel of Fig. 2i) were also increased in *polo*⁹ mutant brains compared to wild type (upper panels of Fig. 2g–i). Concomitantly, a dramatic reduction of differentiated cells expressing neuronal markers, Embryonic Lethal Abnormal Vision (Elav) (lower panel of Fig. 2e) or nuclear Prospero (Pros) (lower panel of Fig. 2f), was observed in *polo*⁹ mutant brains. A similar neuroblast overproliferation phenotype was observed in *polo*¹⁰ and in *trans*-heterozygotes between *polo*⁹ and a deficiency that deletes *polo* (data not shown). A *Polo*-GFP genomic construct fully rescued this *polo* mutant phenotype (Supplementary Fig. 5c), verifying that these defects are caused by *polo* loss-of-function. Excess self-renewal and proliferation at the expense of neuronal differentiation was also observed in MARCM (mosaic analysis with a repressible cell marker) clones derived from single *polo*⁹ mutant neuroblasts (lower panels of Fig. 2j, k, and data not shown). These results indicate that Polo behaves like a tumour suppressor to inhibit neuroblast self-renewal and to promote differentiation. *polo* mutant GMCs may revert to neuroblast-like cells, as has been shown for *brat* (brain tumor) mutants⁵.

We then analysed the physiological role of Polo in regulating Pon localization and function. Most larval neuroblasts were found at metaphase in *polo*⁹ mutant brains, and both Pon (Fig. 3a; $n = 37$, 79%) and Numb (Fig. 3c; $n = 32$, 87%) were uniformly distributed. In late anaphase/telophase neuroblasts, Pon (Fig. 3b; $n = 12$, 58.3%) and Numb (Fig. 3d; $n = 14$, 57.1%) were mis-segregated to both daughter cells. Defective Pon and Numb localization in the *polo*

mutant is unlikely to be a secondary consequence of cell cycle arrest, because arrest of wild-type neuroblasts at metaphase with microtubule-depolymerizing drugs actually increased the number of cells possessing a Numb crescent¹¹.

To test whether Polo is specifically required for Pon/Numb localization, we analysed other apical and basal proteins. In *polo*⁹ mutant neuroblasts, the basal localization of Brat (Fig. 3e; $n = 54$, 79.6%; Supplementary Fig. 4d) and Pros (Supplementary Fig. 4a) was relatively normal. Moreover, double-labelling of the same mutant neuroblast showed that the localization of Mira, the adaptor protein for Pros and Brat, was normal, whereas Pon localization was abnormal (Supplementary Fig. 4b). Introduction of a *Polo*-GFP transgene into *polo*⁹ effectively rescued the Pon localization and cell-cycle defects (Supplementary Fig. 5a). Apical proteins such as Insc, Baz, Pins and Discs Large 1 (Dlg1) were localized normally in *polo*⁹ mutant neuroblasts (Fig. 3f and Supplementary Fig. 4a, c). The only apical protein we found showing abnormal localization was aPKC (Fig. 3g, h, and Supplementary Fig. 4e), which became distributed uniformly on the cortex and showed cytoplasmic localization (Fig. 3g; $n = 38$, 60.5%). During telophase, aPKC could be mis-segregated into both daughter cells (Fig. 3h; $n = 18$, 44.4%).

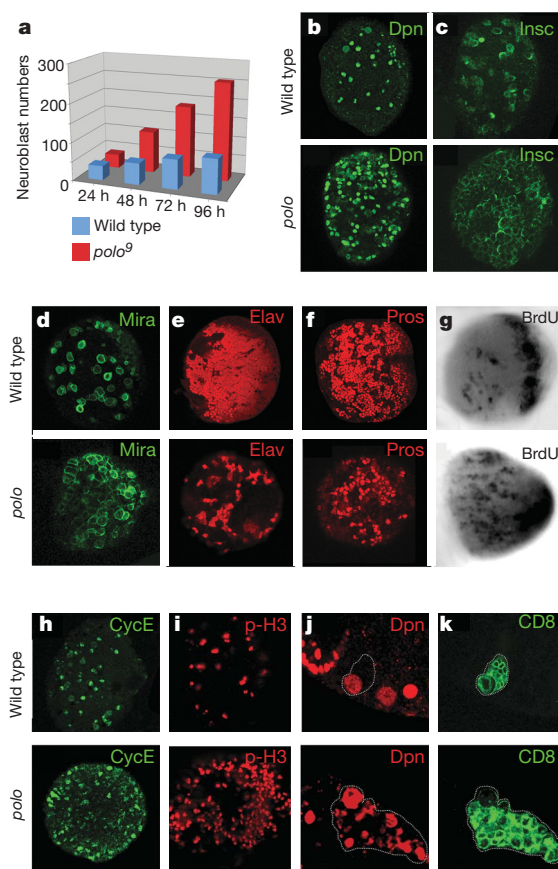


Figure 2 | Polo functions as a tumour suppressor to inhibit neuroblast self-renewal and to promote neuronal differentiation. **a**, Quantification of central brain neuroblast numbers in wild-type and *polo*⁹ mutants from 24- to 96-h ALH. $n = 20$ per time point per genotype. **b–i**, Confocal single-scanning images of wild-type (**b–i**, upper panels) and the *polo*⁹ mutant (**b–i**, lower panels) larval brains at 96 h ALH that are stained with neuroblast markers Dpn (**b**), Insc (**c**) or Mira (**d**), neuronal markers Elav (**e**) or Pros (**f**), or cell proliferation markers BrdU (**g**), CycE (**h**) or phospho-histone H3 (**i**). **j, k**, *polo*⁹ MARCM clones showed the neuroblast overgrowth phenotype. Single neuroblast clones marked by CD8-GFP (and demarcated by stippled line) were generated for the wild-type (**j, k**, upper panels) and the *polo*⁹ mutant (**j, k**, lower panels) in larval brains. Wild-type clones contained one neuroblast that was marked by Dpn (**j**, upper panel; 1.0 ± 0 neuroblast per clone, $n = 9$), whereas *polo*⁹ mutant clones contained multiple neuroblast-like cells (**j**, lower panel; 10.6 ± 3.4 (s.d.) neuroblasts per clone, $n = 7$).

Polo is localized to centrosomes and is required for centrosome organization and separation. We tested whether Polo has a role in orienting neuroblast mitotic spindles. The tight coupling of spindle orientation with crescent formation seen in wild-type neuroblasts (Fig. 3i, j; $n = 24$, 100%) was disrupted in *polo*⁹ metaphase neuroblasts with two centrosomes (Fig. 3k, l; $n = 55$, 56%). Therefore, Polo kinase is also required for coupling mitotic spindle orientation with cortical protein asymmetry. This spindle phenotype was fully rescued by the *Polo*-GFP transgene (Supplementary Fig. 5b).

We next probed the role of Pon phosphorylation in mediating Numb localization. Full-length Pon containing the S611A or S611D mutation was used to assess the effects of S611 phosphorylation. In *pon* mutant neuroblasts, Numb localization was defective (Fig. 4a), as reported previously¹⁶. Introducing wild-type Pon restored Numb asymmetric localization at metaphase and later stages (Fig. 4b; $n = 27$, 100%). Most *pon* mutant neuroblasts expressing Pon(S611D) also showed rescue (Fig. 4c; $n = 12$, 73%). In contrast, *pon* mutant neuroblasts expressing Pon(S611A) showed largely abnormal Numb localization (Fig. 4d; $n = 19$, 86%). Polo-mediated phosphorylation is therefore important for Pon to localize Numb. We also analysed the function of Pon in neuroblast self-renewal by generating *pon* MARCM clones. Interestingly, compared to wild-type clones (Fig. 4e, f), ectopic neuroblast self-renewal was observed in *pon* mutant clones (Fig. 4g, h).

We have shown that *polo* mutations affect Numb and aPKC localization as well as spindle orientation—processes known to affect neuroblast self-renewal to various degrees^{8,9,18–20}. Strikingly, overexpression of either Numb-GFP or Numb effectively suppressed

the ectopic neuroblast self-renewal phenotype seen in the *polo*⁹ mutant (Fig. 4i–k and data not shown). This effect was not caused by increased neuroblast apoptosis (Supplementary Fig. 6), and overexpression of Numb-GFP or Numb in a wild-type background did not affect the neuroblast number (Fig. 4k, Supplementary Fig. 6 and

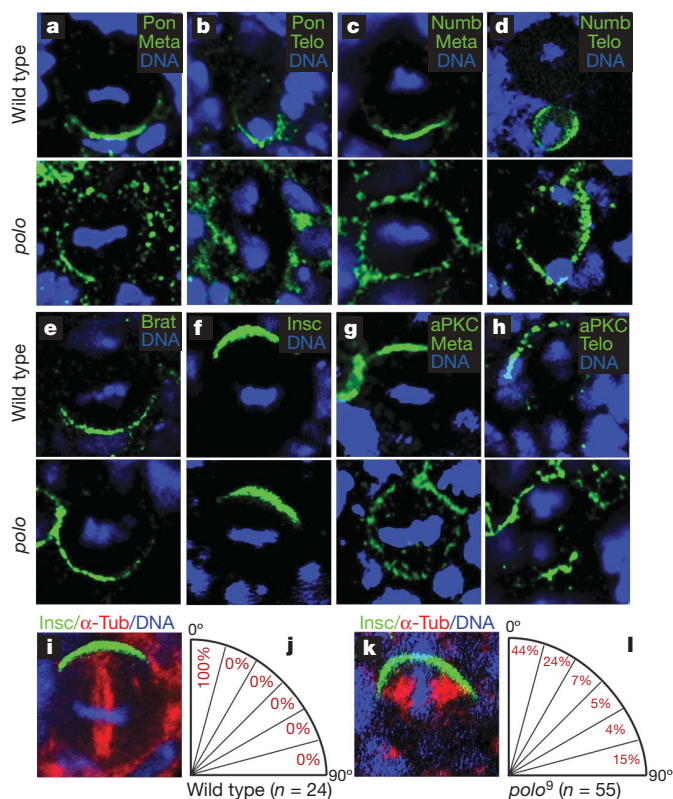


Figure 3 | Polo kinase is required for the asymmetric localization of Pon and Numb and for correct spindle orientation. Neuroblasts from wild type (upper panels) and the *polo*⁹ mutant (lower panels) were immunostained for Pon (a, b), Numb (c, d), Brat (e), Insc (f) and aPKC (g, h). The neuroblasts shown in b, d and h were at telophase. i–l, Abnormal spindle orientation phenotype in the *polo* mutant. Wild-type (i) and *polo*⁹ mutant (k) neuroblasts were triple-labelled for Insc, α -tubulin (α -Tub) and DNA. Spindle orientation in wild-type ($n = 22$) and *polo*⁹ mutant ($n = 53$) neuroblasts was quantified in j and l, respectively.

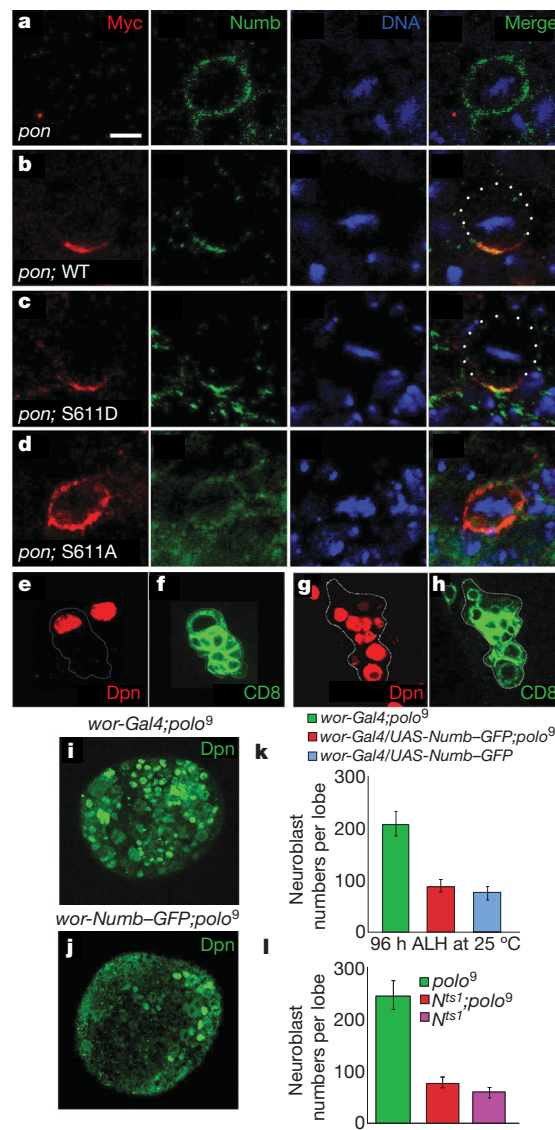


Figure 4 | Polo controls Numb asymmetry through phosphorylation of Pon, and neuroblast self-renewal through regulation of Numb. a–d, Phosphorylation of Pon S611 mediates the asymmetric localization of Numb. Numb exhibits uniform cortical localization in *pon* mutant metaphase neuroblasts (a). The Numb localization defect in *pon* mutant neuroblasts was rescued by wild-type Pon (b) and Pon(S611D) (c), but not by Pon(S611A) (d). e–h, In wild-type (e, f) and *pon* (g, h) MARCM clones marked by CD8-GFP and demarcated by stippled lines, multiple neuroblasts (marked by Dpn) were observed in clones derived from single *pon* mutant neuroblasts (g), whereas only one neuroblast was observed in wild-type clones (e). i–k, Overexpression of Numb, using *wor-Gal4* to drive expression of UAS-Numb-GFP (*wor-Numb-GFP*), suppressed the supernumerary neuroblast phenotype in the *polo*⁹ mutant. Neuroblasts were marked by Dpn (green) in *wor-Gal4;polo*⁹ (i) and *wor-Numb-GFP;polo*⁹ (j) brains. Results from overexpressing *wor-Numb* are shown in Supplementary Fig. 7. k, Quantitative analysis of central brain neuroblast numbers for *wor-Gal4;polo*⁹, *wor-Numb-GFP;polo*⁹ and *wor-Numb-GFP* 96 h ALH at 25 °C ($n = 20$). l, Loss of function of Notch suppresses neuroblast overgrowth phenotype of *polo* mutant. Quantification of central brain neuroblast numbers for *polo*⁹, *Notch^{ts1};polo*⁹ (*Nts1;polo*⁹) and *Notch^{ts1}* (*Nts1*) was performed 72 h ALH at 29 °C ($n = 20$). k, l, Error bars are s.d.

data not shown). These results indicate that Numb is a principal player downstream of Polo in regulating neuroblast self-renewal. Numb overexpression did not rescue the aPKC mislocalization and spindle misorientation phenotypes of *polo* mutants (Supplementary Fig. 7). These defects could also contribute to the neuroblast overgrowth phenotype of *polo* mutants, but their effects might have been masked by Numb overexpression. Consistent with this, introduction of Pon(S611D) into a *polo* mutant neuroblast did not significantly rescue the neuroblast overgrowth phenotype, despite the partial restoration of Numb localization (Supplementary Fig. 8). Because aPKC localization and spindle orientation defects were not rescued by Pon(S611D) (Supplementary Fig. 8), these defects may account for the inability of Pon(S611D) to rescue the overgrowth phenotype of *polo*. aPKC has been shown to phosphorylate Numb (ref. 21). Delocalized aPKC at the basal side may be sufficient to inactivate endogenous Numb, but not overexpressed Numb, owing to titration by the overexpressed protein.

Numb was previously shown to inhibit neuroblast self-renewal by antagonizing Notch signalling⁹. Reducing *Notch* significantly suppressed the neuroblast overgrowth phenotype of the *polo*⁹ mutant (Fig. 4l and Supplementary Fig. 9). Reducing *Notch* in a wild-type background also led to a partial loss of neuroblasts (Fig. 4l and Supplementary Fig. 9), consistent with Notch being required for progenitor self-renewal. We envision that in *polo* or *pon* mutants, owing to the symmetric distribution of Numb, the GMCs receive insufficient Numb to inhibit Notch, thereby causing them to adopt a neuroblast-like fate. To test further the importance of Numb asymmetric localization in neuroblast self-renewal versus differentiation, we analysed the *numb*^{S52F} mutation, which apparently affects Numb asymmetric localization but not its stability or activity²². In *numb*^{S52F} neuroblast clones, ectopic neuroblast self-renewal similar to that seen in *polo* or *pon* clones was observed (Supplementary Fig. 10). Thus, loss of Numb asymmetric localization is sufficient to cause neuroblast overgrowth.

Our results indicate that Polo controls the self-renewal versus differentiation decision of neural progenitors by regulating the localization and activity of Numb and the orientation of mitotic spindles. Polo regulates the localization of Numb by means of Pon. Although immunofluorescence shows that Polo is primarily localized to the centrosomes, a cytosolic pool of Polo probably exists because Polo exhibits dynamic *in vivo* association with the mitotic apparatus and many non-centrosomal Polo substrates have been identified^{15,23}. Cytosolic localization of the centrosomal kinase Aurora-A has been demonstrated²⁴. How Polo regulates the localization of aPKC, the activity of Numb and the orientation of mitotic spindles awaits further investigation. In addition to the Numb/Notch pathway, other factors such as Pros and Brat are necessary for preventing GMCs from undergoing self-renewing divisions^{4–6,25}. Because these factors are segregated normally in *polo* neuroblasts, it seems that they are not sufficient to prevent progenitor self-renewal or that activation of Notch is able to override their effects. Intriguingly, some Plks behave as tumour suppressors in mammals, and loss of Numb has also been implicated in the hyperactivation of Notch signalling in breast cancer cells^{15,26,27}. Our results and future studies in models like *Drosophila* will provide mechanistic insights into these observed tumour-suppressor roles of Polo and Numb.

METHODS SUMMARY

Fly genetics. Fly culture and crosses were performed according to standard procedures and were raised at indicated temperatures. To generate Pon(S611A) and Pon(S611D) transgenic flies, the corresponding complementary DNA constructs generated in pUAST vector were injected into *w*⁻ embryos to obtain transgenic lines according to established procedures. All other fly stocks and *Gal4* lines were obtained from the Bloomington *Drosophila* stock centre or as described^{17,23,28,29}.

Antibody production and immunohistochemistry. A synthetic peptide with the following sequence CLADVPsIIDTSE was conjugated to the keyhole limpet

haemocyanin carrier protein and used to elicit anti-p-S611 antibody production at Open Biosystems. Immunohistochemistry on embryos or larval brain tissues was performed as described¹⁶. The primary antibodies used were: anti-myc (1:200), anti-GFP (1:2,000), anti-p-S611 (1:20), anti-Pon (1:1,000), anti-Elav (1:20), anti-CycE (1:10), anti-Dpn (1:1,000), anti-Mira (1:1,000), anti-Numb (1:1,000), anti-Baz (1:500), anti-Insc (1:500), anti-Pins (1:500), anti-Dlg (1:200), anti-aPKC (1:500), anti-Brat (1:100), anti-Pros (1:20) and anti-p-histone (1:500).

In vitro phosphorylation assays. The conditions of the *in vitro* kinase assays were essentially as described³⁰.

Clonal analysis. To generate single neuroblast clones in MARCM analysis, 24 h ALH larvae were heat-shocked at 37 °C for 90 min and further aged for 3 days at 25 °C.

BrdU labelling. Dissected larval tissue was given a 40-min pulse of 37.5 µg ml⁻¹ BrdU in Shields and Sang 3 M insect medium, and fixed for 15 min in 3.7% formaldehyde. DNA was denatured with 2 M HCl for 40 min, and tissue was washed with PBS and incubated with anti-BrdU.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 May; accepted 28 June 2007.

- Morrison, S. J. & Kimble, J. Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature* **441**, 1068–1074 (2006).
- Jan, Y. N. & Jan, L. Y. Asymmetric cell division in the *Drosophila* nervous system. *Nature Rev. Neurosci.* **2**, 772–779 (2001).
- Causinus, E. & Gonzalez, C. Induction of tumor growth by altered stem-cell asymmetric division in *Drosophila melanogaster*. *Nature Genet.* **37**, 1125–1129 (2005).
- Betschinger, J., Mechtler, K. & Knoblich, J. A. Asymmetric segregation of the tumor suppressor Brat regulates self-renewal in *Drosophila* neural stem cells. *Cell* **124**, 1241–1253 (2006).
- Lee, C. Y., Wilkinson, B. D., Siegrist, S. E., Wharton, R. P. & Doe, C. Q. Brat is a Miranda cargo protein that promotes neuronal differentiation and inhibits neuroblast self-renewal. *Dev. Cell* **10**, 441–449 (2006).
- Bello, B., Reichert, H. & Hirth, F. The brain tumor gene negatively regulates neural progenitor cell proliferation in the larval central brain of *Drosophila*. *Development* **133**, 2639–2648 (2006).
- Rhyu, M. S., Jan, L. Y. & Jan, Y. N. Asymmetric distribution of numb protein during division of the sensory organ precursor cell confers distinct fates to daughter cells. *Cell* **76**, 477–491 (1994).
- Lee, C. Y. *et al.* *Drosophila* Aurora-A kinase inhibits neuroblast self-renewal by regulating aPKC/Numb cortical polarity and spindle orientation. *Genes Dev.* **20**, 3464–3474 (2006).
- Wang, H. *et al.* Aurora-A acts as a tumor suppressor and regulates self-renewal of *Drosophila* neuroblasts. *Genes Dev.* **20**, 3453–3463 (2006).
- Tio, M., Udolph, G., Yang, X. & Chia, W. *cdc2* links the *Drosophila* cell cycle and asymmetric division machineries. *Nature* **409**, 1063–1067 (2001).
- Knoblich, J. A., Jan, L. Y. & Jan, Y. N. Asymmetric segregation of Numb and Prospero during cell division. *Nature* **377**, 624–627 (1995).
- Zhou, Y. *et al.* The mammalian Golgi regulates numb signaling in asymmetric cell division by releasing ACBD3 during mitosis. *Cell* **129**, 163–178 (2007).
- Sunkel, C. E. & Glover, D. M. *polo*, a mitotic mutant of *Drosophila* displaying abnormal spindle poles. *J. Cell Sci.* **89**, 25–38 (1988).
- Barr, F. A., Silje, H. H. & Nigg, E. A. Polo-like kinases and the orchestration of cell division. *Nature Rev. Mol. Cell Biol.* **5**, 429–440 (2004).
- van de Weert, B. C. & Medema, R. H. Polo-like kinases: a team in control of the division. *Cell Cycle* **5**, 853–864 (2006).
- Lu, B., Rothenberg, M., Jan, L. Y. & Jan, Y. N. Partner of Numb colocalizes with Numb during mitosis and directs Numb asymmetric localization in *Drosophila* neural and muscle progenitors. *Cell* **95**, 225–235 (1998).
- Lu, B., Ackerman, L., Jan, L. Y. & Jan, Y. N. Modes of protein movement that lead to the asymmetric localization of partner of Numb during *Drosophila* neuroblast division. *Mol. Cell* **4**, 883–891 (1999).
- Lee, C. Y., Robinson, K. J. & Doe, C. Q. Lgl, Pins and aPKC regulate neuroblast self-renewal versus differentiation. *Nature* **439**, 594–598 (2006).
- Izumi, Y., Ohta, N., Hisata, K., Raabe, T. & Matsuzaki, F. *Drosophila* Pins-binding protein Mud regulates spindle-polarity coupling and centrosome organization. *Nature Cell Biol.* **8**, 586–593 (2006).
- Siller, K. H., Cabernard, C. & Doe, C. Q. The NuMA-related Mud protein binds Pins and regulates spindle orientation in *Drosophila* neuroblasts. *Nature Cell Biol.* **8**, 594–600 (2006).
- Smith, C. A. *et al.* aPKC-mediated phosphorylation regulates asymmetric membrane localization of the cell fate determinant Numb. *EMBO J.* **26**, 468–480 (2007).
- Bhalerao, S., Berdnik, D., Torok, T. & Knoblich, J. A. Localization-dependent and -independent roles of numb contribute to cell-fate specification in *Drosophila*. *Curr. Biol.* **15**, 1583–1590 (2005).

23. Moutinho-Santos, T., Sampaio, P., Amorim, I., Costa, M. & Sunkel, C. E. *In vivo* localisation of the mitotic POLO kinase shows a highly dynamic association with the mitotic apparatus during early embryogenesis in *Drosophila*. *Biol. Cell* **91**, 585–596 (1999).
24. Berdnik, D. & Knoblich, J. A. *Drosophila* Aurora-A is required for centrosome maturation and actin-dependent asymmetric protein localization during mitosis. *Curr. Biol.* **12**, 640–647 (2002).
25. Choksi, S. P. *et al.* Prospero acts as a binary switch between self-renewal and differentiation in *Drosophila* neural stem cells. *Dev. Cell* **11**, 775–789 (2006).
26. Pece, S. *et al.* Loss of negative regulation by Numb over Notch is relevant to human breast carcinogenesis. *J. Cell Biol.* **167**, 215–221 (2004).
27. Stylianou, S., Clarke, R. B. & Brennan, K. Aberrant activation of Notch signaling in human breast cancer. *Cancer Res.* **66**, 1517–1525 (2006).
28. Clarke, A. S., Tang, T. T., Ooi, D. L. & Orr-Weaver, T. L. POLO kinase regulates the *Drosophila* centromere cohesion protein MEI-S332. *Dev. Cell* **8**, 53–64 (2005).
29. Ohshiro, T., Yagami, T., Zhang, C. & Matsuzaki, F. Role of cortical tumour-suppressor proteins in asymmetric division of *Drosophila* neuroblast. *Nature* **408**, 593–596 (2000).
30. Nishimura, I., Yang, Y. & Lu, B. PAR-1 kinase plays an initiator role in a temporally ordered phosphorylation process that confers tau toxicity in *Drosophila*. *Cell* **116**, 671–682 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Glover, F. Matsuzaki, Y. N. Jan, A. Wodarz, C. Doe, J. Knoblich, L. Luo, T. Orr-Weaver, C. Sunkel, B. Edgar, H. Richardson, F. Schweisguth, J. B. Skeath, A. Gould, E. D. Schejter, the Developmental Studies Hybridoma Bank and the Bloomington Stock Center for fly stocks and antibodies. We thank S. Guo for reading the manuscript, and F. Yu and members of the Lu and Chia laboratories for discussion and help. H.W. thanks U. Heberlein for support and S. G. S. Ling for technical help. This was supported by a NIH grant (B.L.) and Temasek Lifesciences Laboratory funding (W.C.)

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.L. (bingwei@stanford.edu) or W.C. (wchia@tll.org.sg).

METHODS

Immunohistochemistry. Larval brains were dissected in PBS, and were fixed in PBS containing 3.7% formaldehyde for 15 min at room temperature (22–25 °C). Brains were washed several times with PBS containing 0.1% Triton X-100 (PBT) and blocked in PBT with 3% BSA for 30 min. After blocking, brains were incubated with the appropriate primary antibody in PBT overnight at 4 °C. After washing three times with PBT for 10 min each time, brains were incubated with secondary antibodies for 2 h, followed by three washes with PBT for 10 min each time. For DNA staining, TO-PRO-3 (Molecular Probes) was added in the final wash at a dilution of 1:20,000. Brains were then mounted on slides in Vectashield (Vector Labs) and were ready for imaging.

For embryo staining, stage 9 and 10 embryos were collected and fixed with 4% formaldehyde:heptane (1:1 volume) at room temperature for 30 min and processed as described for larval-brain staining. For p-S611-Pon antibody staining, the larval brains were dissected in Schneider's medium (Sigma), fixed with 4% formaldehyde in PEM buffer (100 mM Pipes, pH 7.0, 1 mM EGTA, 1 mM MgCl₂) at 4 °C for 2–3 h, and then blocked for 30 min in PBST (PBS plus 0.2% Triton X-100) containing 10% NGS (normal goat serum, JacksonImmuno Research). After blocking, samples were incubated with primary antibody (1:20) in PBST plus 10% NGS overnight at 4 °C. After primary antibody incubation, samples were briefly rinsed once with PBST and were washed three times in PBST for 30 min each. The secondary antibody used was biotinylated goat anti-rabbit, which was detected with fluorescently labelled streptavidin. The specimens were extensively washed in PBST before mounting in hardset antifade mounting medium (Vectorshield H-1400, Vector Labs).

Neuroblast quantification and brain orientation. Quantifications of larval central brain neuroblast numbers were done on samples 24 h to 96 h ALH at 25 °C. For each genotype, 4–6 h embryo collections were obtained from a bottle with 100–200 flies. Embryos were allowed to develop for 20 h at 25 °C before hatching. Thereafter, time course experiments were performed 24 h, 48 h, 72 h and 96 h ALH according to the midpoint of the 4–6 h time window. Around 15–20 larvae from each time point were dissected, were stained with anti-Dpn antibody (1:1,000), and immunohistochemistry was performed as in the above paragraph. Central brain neuroblasts can be distinguished from optic lobe neuroblasts on the basis of their medial-superficial location in the brain, larger size and dispersed pattern. Around 20 brain lobes were quantified for each genotype. Larval brains were mounted with their dorsal side up. The midline between the two brain lobes is to the left in all panels.

Spindle-orientation quantification. Larval neuroblasts were stained with rat anti-Tubulin (1:250, AbCam) and rabbit anti-Insc or rabbit anti-CNN (1:5,000) and mouse anti-Mira (1:50). Confocal images were captured and were used for quantification. One line was drawn parallel to the metaphase spindle (indicated by tubulin or CNN signals) and another line perpendicular to the

apical crescent (marked by Insc) or the basal crescent (marked by Mira). The angle at which the two lines met was then measured.

BrdU labelling. The larvae were aged and the brains dissected in Shields and Sang M3 insect media supplemented with 5% NGS. The brains were incubated for 40 min at room temperature in the Shields and Sang media plus NGS with 37.5 µg ml⁻¹ BrdU. They were then washed in PBT and fixed in PBT containing 4% formaldehyde for 15 min, and were washed again in PBT before incubation with 2 M HCl for 40 min. The brains were washed repeatedly, and then blocked in PBT plus NGS. Standard immunohistochemical detection was then performed to detect the BrdU with 1:20 mouse anti-BrdU (Roche).

Analysis on *Notch*^{ts1}. *Notch*^{ts1} flies were kept at 19 °C for 8 h for embryo collection. Embryos were allowed to develop at 19 °C for 40 h before hatching. Hatched larvae were then immediately shifted to 30 °C for 72 h before analysis.

Constructs. All the DNA point mutations were introduced using QuickChange II XL site-directed mutagenesis kit (Stratagene). The mutations in pEGFP-Pon-LD and pcDNA3-Pon-FL (full length) were screened and confirmed by DNA sequencing before subcloning into pUAST vectors. Primer sequences are available on request. Transgenic flies of pUAST-GFP-Pon-LD mutations were generated in our laboratory, and pUAST-Pon-FL S611A and S611D transgenic flies were generated by BestGene.

Purification of Pon-LD wild-type and S611A proteins. The DNA insertion encoding Pon-LD (493–672 amino acids) was fused with His tag at the C terminus and cloned into pGEX6P-1 vectors. Pon-LD containing S611A mutations were amplified by PCR using pEGFP-Pon(S611A)-LD as template and were cloned into pGEX6P-1 vectors. Wild type and S611A in these expression constructs were confirmed by sequencing before they were transformed into expression strains Rosetta2 or Tuner (Novagen). The expression strains containing the desired wild-type and S611A constructs were cultured in Terrific Broth medium (Invitrogen) at 37 °C for 2–3 h and then induced for expression with the addition of 0.5 mM IPTG at 30 °C for 6–7 h. The cell pellets were suspended in TBS buffer (20 mM Tris, pH 8.0, 200 mM NaCl) and lysed with 0.5% Triton X-100. Wild-type and mutant proteins were purified using Ni-NTA column (Qiagen) and eluted from the column with 75 mM imidazole. The fractions were further concentrated with Centricon (Millipore) and stored at –80 °C in aliquots.

In vitro phosphorylation assay. 0.5–1 µg wild-type and S611A proteins were incubated with 2 µl human Polo kinase1 (hPLK1, Cell Signaling) in a kinase buffer containing 50 mM Tris, pH 7.2, 15 mM MgCl₂, 100 µM ATP, 5 mM β-glycerophosphate (Sigma). The phosphorylation reaction was extended for 30 min at 30 °C with the addition of [γ -³²P] hot ATP (3 µCi per sample, PerkinElmer). SDS sample buffer was added into the mixture to stop the kinase reaction. Samples were run with SDS-PAGE and were subjected to autoradiography.

Fc receptor but not complement binding is important in antibody protection against HIV

Ann J. Hessel^{1*}, Lars Hangartner^{1*}, Meredith Hunter², Carin E. G. Havenith³, Frank J. Beurskens³, Joost M. Bakker³, Caroline M. S. Lanigan¹, Gary Landucci⁴, Donald N. Forthal⁴, Paul W. H. I. Parren³, Preston A. Marx² & Dennis R. Burton¹

Most successful vaccines elicit neutralizing antibodies and this property is a high priority when developing an HIV vaccine^{1,2}. Indeed, passively administered neutralizing antibodies have been shown to protect against HIV challenge in some of the best available animal models. For example, antibodies given intravenously can protect macaques against intravenous or mucosal SHIV (an HIV/SIV chimera) challenge and topically applied antibodies can protect macaques against vaginal SHIV challenge^{3,4}. However, the mechanism(s) by which neutralizing antibodies afford protection against HIV is not understood and, in particular, the role of antibody Fc-mediated effector functions is unclear. Here we report that there is a dramatic decrease in the ability of a broadly neutralizing antibody to protect macaques against SHIV challenge when Fc receptor and complement-binding activities are engineered out of the antibody. No loss of antibody protective activity is associated with the elimination of complement binding alone. Our *in vivo* results are consistent with *in vitro* assays indicating that interaction of Fc-receptor-bearing effector cells with antibody-complexed infected cells is important in reducing virus yield from infected cells. Overall, the data suggest the potential importance of activity against both infected cells and free virus for effective protection against HIV.

Neutralization is the most sought-after property of vaccine-induced antibodies and is a property that is measured *in vitro* as the ability of antibodies to inhibit viral entry into target cells in the absence of other anti-viral functions. Generally, neutralization results from inhibition of viral attachment, fusion events, or both^{5,6}. *In vivo*, neutralizing antibodies can also exert anti-viral effects by other mechanisms, including effector functions mediated by the crystallizable fragment (Fc) part of the antibody molecule, such as complement activation and antibody-dependent cellular cytotoxicity (ADCC). These effector functions can act against both free virions and virus-infected cells⁷.

The critical importance of effector functions for antibody-mediated protection against a number of viruses has been well documented^{5,8,9}, but there are few reports for HIV or SIV. In post-exposure studies of HIV-1-challenged hu-PBL SCID mice, it was reported¹⁰ that treatment of mice with cobra-venom-factor to inactivate serum complement activity abrogated protection. In established SIV infection in macaques, polyclonal non-neutralizing anti-SIV immunoglobulin G was passively infused and it was suggested that ADCC might be responsible for producing a modest decline in levels of SIV-infected cells¹¹. Serum ADCC activity measured *in vitro* has also been associated with protection^{12,13}, although

other studies have not found a correlation¹⁴. ADCC can be triggered by non-neutralizing antibodies^{15,16}, presumably by binding to non-functional forms of envelope on the surface of infected cells.

Here we explore the role of antibody effector function in protection against HIV-1 in the SHIV/macaque model using a neutralizing human monoclonal antibody and variants of this antibody in which effector functions have been specifically disabled. The human monoclonal antibody b12 has been shown to neutralize a wide range of HIV isolates from different clades^{17,18} and, when administered systemically or topically, to protect macaques against vaginal challenge with a SHIV that uses the chemokine receptor 5 (CCR5)^{4,19}. The antibody recognizes a conserved epitope overlapping the CD4-binding site of gp120 (ref. 20). This epitope constitutes a major target of HIV vaccine design.

The generation and characterization of a series of Fc variants of antibody b12 has been described²¹. Here we focused on two antibody variants (Fig. 1a) from that study. The K322A (KA) variant, designed to abrogate complement activation, was shown to bind human complement C1q with greatly lowered affinity and to inefficiently activate human complement. The L234A,L235A (LALA) variant was shown to bind human Fcγ receptors weakly, to trigger ADCC with lowered efficiency and also to bind and activate human complement weakly. Figure 1b, c confirms, as predicted and as described in part in ref. 21, the antigen-binding and neutralization properties of the variant antibodies are equivalent to those of wild-type b12. Next we compared the abilities of the antibodies to interact with human and macaque effector-triggering molecules. Both variants show markedly decreased binding of human and macaque C1q (Fig. 2a) and complement activation as assessed by C3 fixation (Fig. 2b). Thus, neither variant activates the classical pathway of complement, and the absence of C3 fixation also rules out activation by the alternative or lectin pathways. Figure 2c shows that b12 and the KA variant interact similarly with recombinant human and macaque FcγRIa, FcγRIIa and FcγRIIIa, and with human FcγRIIb, but the LALA variant interacts with FcγRs considerably less well. Of note, the LALA variant is defective in binding to both FcγRIIa and FcγRIIb so major effects arising from disproportionate signalling by activating and inhibitory receptors²² are not anticipated. In summary, wild-type b12 is complement and FcγR competent, the KA variant is disabled with respect to complement but competent with respect to FcγR binding, and the LALA variant is defective both in terms of complement and FcγR binding.

We next investigated the effect of Fc mutation on pharmacokinetics of the three b12 antibodies in macaques. None of the

¹Departments of Immunology and Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, USA. ²Tulane National Primate Research Center, Tulane University, Covington, Louisiana 70433, USA. ³Genmab, 3584 CM Utrecht, The Netherlands. ⁴Division of Infectious Diseases, Department of Medicine, University of California, Irvine School of Medicine, Irvine, California 92697, USA.

*These authors contributed equally to this work.

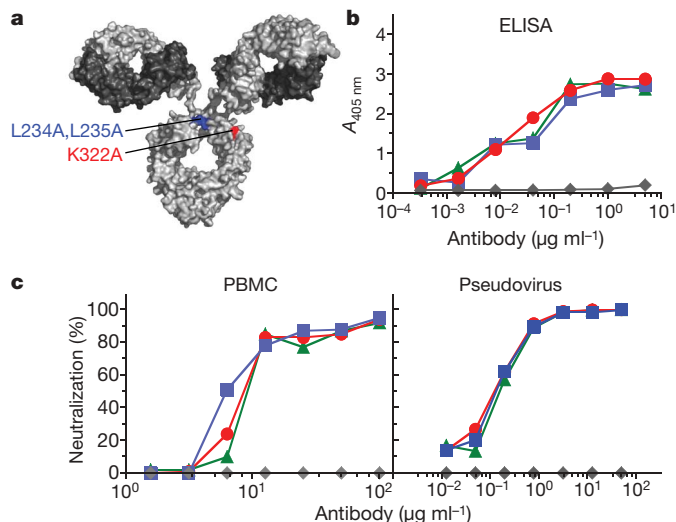


Figure 1 | Effector function variants of b12: location of substitutions, antigen binding and neutralization properties. **a**, The molecular surface representation of IgG1 b12 depicts the location of the substitutions K322A and L234A, L235A. **b**, Binding, by ELISA, of b12 and variants to SHIV_{SF162P3} gp120 captured from a viral lysate. **c**, Two assay formats of SHIV_{SF162P3} neutralization by b12 and variants: PBMC-based neutralization and a single-round replication luciferase reporter assay of pseudotyped virus. Controls for ELISA and PBMC-based neutralization are non-HIV human IgG1 mAbs; the control for the pseudovirus assay is serum from an HIV-1 seropositive patient (N16). Image provided by E. O. Saphire. Symbols as in Fig. 2.

mutations is close to the neonatal Fc receptor (FcRn)-binding site responsible for the maintenance of IgG half-lives, and indeed the half-lives of all three b12 species in macaques were similar (Supplementary Table 1).

A protection experiment (Fig. 3), comparing the abilities of the three b12 antibodies to prevent or modulate infection by a CCR5-using SHIV, was performed following Depo-Provera treatment of all animals to thin the vaginal epithelium¹⁹. One day before vaginal challenge with 300 50% tissue culture infectious doses (TCID₅₀) of SHIV_{SF162P3}, each animal was given an intravenous dose of 25 mg kg⁻¹ of one of the b12 antibodies or an isotype control human IgG. The experiment included nine animals in each of the b12 and b12 variant groups and four animals in the control group. Blood was drawn frequently to monitor viral infection, transferred antibody levels and serum-neutralizing activity. All 4 controls became infected with a peak viraemia of 3×10^6 – 4×10^7 virus copies per ml between days 14 and 21. In the b12 group, 8 out of 9 animals were protected and showed no detectable virus at day 159 following challenge. The single infected animal showed a relatively low peak viraemia of 10^6 virus copies per ml at a time point (day 38) much delayed when compared to controls. Similarly, in the KA variant group, 8 out of 9 animals were protected, with a single infected animal showing a relatively low peak viraemia (8×10^5 virus copies per ml) occurring late (day 42) after challenge. The viral loads in the control group differed significantly from those in both the b12 and the KA variant groups from day 10 post challenge and onwards ($P < 0.01$, two-way analysis of variance (ANOVA) with Bonferroni post hoc test; see Supplementary Table 3 for detailed day-by-day analyses.)

In contrast, transfer of the LALA variant left 4 out of 9 animals unprotected with similar peak viraemia to that of control animals, in

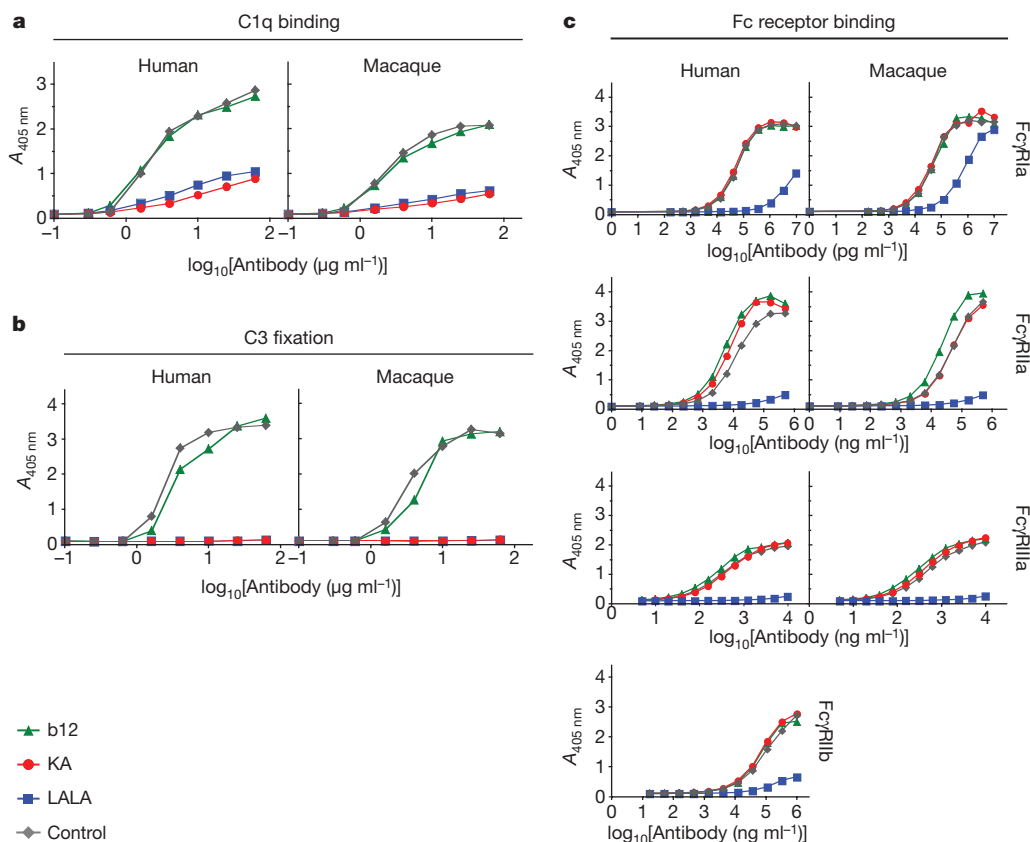


Figure 2 | Interaction of b12 and variants with effector molecules. **a**, Binding of b12 and variants to C1q from human and macaque serum. **b**, C3 fixation by b12 and variants in human and macaque serum. Wild-type b12 and IgG1 isotype control bind C1q and fix C3. KA and LALA variants show little C1q binding and little C3 fixation. **c**, Binding of b12 and variants to

human and macaque activating FcγRs Ia, IIa and IIIa and to human inhibitory FcγRIIb. The KA variant shows similar FcγR binding, whereas the LALA variant shows weak or no binding to human or macaque FcγRs. (Macaque FcγRIIb is not available.) See Supplementary Table 4 for apparent affinities (IC₅₀).

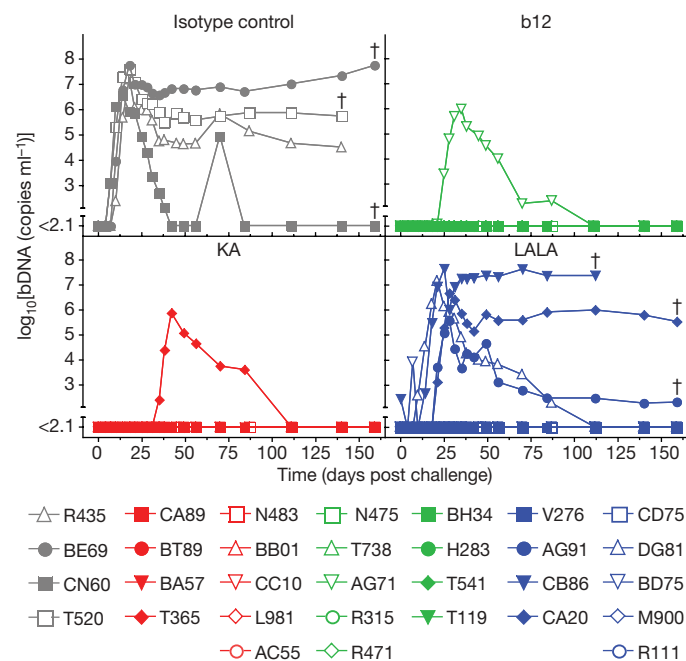


Figure 3 | Temporal analysis of plasma viral loads. All animals in the control group (4/4) experienced peak viraemia (3×10^6 – 4×10^7) between days 14 and 21. Only 1 of 9 animals in both the wild-type b12 and KA variant groups became infected, each with a much-delayed onset of infection and decreased peak viraemia. In contrast, 4 animals became infected in the LALA variant group with viraemia onset and peak similar to the control group. The assay minimum detection limit stated by the manufacturer is 125 bDNA copies per ml ($2.1 \log_{10}$ [bDNA (copies per ml)]) with a 95% confidence level. Protected animals showed no consistently measurable infection. Each curve depicts an individual animal. †, Euthanized animals as a result of virally induced pathology.

the range of 4×10^5 – 1×10^7 virus copies per ml, albeit somewhat delayed. Indeed, two-way ANOVA with Bonferroni post hoc test revealed a difference in viral loads between the control group and the LALA variant group during early infection (day 10, $P < 0.05$; days 14–21, $P < 0.001$; day 25, $P < 0.05$). After day 25, a difference between the control group and the LALA variant group could no longer be observed (except for a difference between both groups on day 70 ($P < 0.01$), which was due to an unusually high viral titre in the control group on that day; see Supplementary Table 3). Taken together, our statistical analyses indicate that the b12 and KA antibodies confer profound protection from viraemia, whereas the LALA antibody confers only partial protection in which infected animals develop peak viraemias similar to controls.

The neutralizing antibody titres (half-maximal inhibitory concentration, IC_{50}) in the sera of the animals at the time of challenge varied between 1:1,592 and 1:4,881 (Supplementary Table 2). Statistical analysis revealed no differences in the neutralization titres at the time of challenge among the groups ($P > 0.05$, see Supplementary Table 3), indicating that differences in the levels of serum-neutralizing activity were not responsible for the patterns of protection described.

To investigate further the status of the protected animals that showed no detectable virus by branched DNA (bDNA) measurements, one protected animal from each of the groups was depleted of $CD8^+$ T cells using a mouse–human chimaeric anti- $CD8$ antibody, a proven procedure to detect virus from monkeys that previously suppressed viral loads below the level of detection²³. None of the animals showed detectable virus following $CD8^+$ T-cell depletion (Supplementary Fig. 1), indicating that sterilizing immunity was probably achieved in the protected animals.

Overall, the challenge experiments reveal that antibody can prevent HIV infection without the recruitment of effector functions. However, in the absence of $Fc\gamma$ -receptor-binding function, the

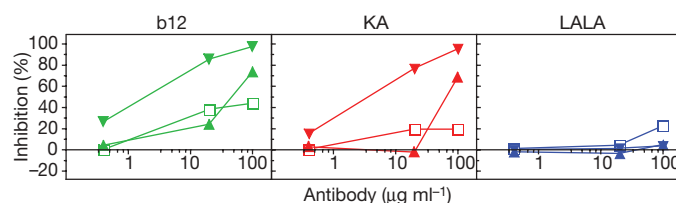


Figure 4 | Antibody-dependent cell-mediated viral inhibition (ADCVI) by b12 and variants. CEM.NKR-CCR5 cells were infected with SHIV_{SF162P3} and incubated for 48 h, washed and combined with effector cells and serially diluted antibody. After a 30-min incubation, cells were re-washed to remove excess antibody and free virus. Viral inhibition was measured after incubation for 7 days. The LALA variant is less effective than wild-type b12 and the KA variant in mediating ADCVI with both human and rhesus effector cells. Squares, no effectors; up-triangles, huPBMCs; down-triangles, rhPBMCs.

likelihood of infection is greatly increased. This raises the questions: (1) how does $Fc\gamma$ -receptor-binding function influence antibody protection, and (2) are there implications for HIV vaccine design? $Fc\gamma$ -receptor-binding may influence protection through clearance of viral particles. Thus, LALA-coated virions, lacking $Fc\gamma$ -receptor-binding, may be less efficiently cleared, for example, by phagocytic cells, than b12-coated virions, and this may increase the chances of infection. Alternatively, the LALA variant may be unable to mediate activities such as ADCC against infected cells and this vulnerability may increase the probability of a productive infection. To investigate the latter possibility, we used an antibody-dependent cell-mediated virus inhibition (ADCVI) assay²⁴ to compare the ability of wild-type and variant b12 antibodies to inhibit virus yield from infected cells in the presence of $Fc\gamma$ -receptor-bearing effector cells *in vitro*. Figure 4 shows that wild-type b12 and the KA variant provide better control of virus in the presence of effector cells than the LALA variant.

A reasonable explanation of the protection experiments described, taking into account the *in vitro* observations made, is as follows. After virus challenge, all infectious virions may be neutralized by antibody and infection of the animal prevented. Alternatively, some cells may become infected and produce large numbers of progeny virus particles. These also may be neutralized and infection prevented. However, as more cells become infected, the risk of infection spinning out of control increases. Lacking $Fc\gamma$ -receptor-binding function, the LALA variant may be less effective in clearing virions and in suppressing viral release from infected cells. The latter function requires higher antibody concentrations than neutralization²¹ (compare Figs 1c and 4) and may contribute to the high concentrations often associated with passive protection against HIV. If activity against infected cells was provided instead by a robust cellular immune response, then considerably lower antibody concentrations may nevertheless provide benefit. In other words, HIV vaccine efficacy may be crucially dependent on a combination of antibody and cell-mediated immunity.

METHODS SUMMARY

Animals used in this study were female Indian rhesus macaques challenged with SHIV_{SF162P3} (refs 25–27). CH variants of b12 were created by site-directed mutagenesis²¹. Recombinant antibodies were expressed in Chinese hamster ovary (CHO-K1) cells, purified using Protein A affinity matrix, and monitored for endotoxin contamination. Plasma viral loads were assessed by quantitative assays for the measurement of SIV RNA using a branched DNA signal amplification assay for SIV²⁸ (Siemens, formerly Bayer). Antibody concentrations in serum and measurement of binding to SHIV_{SF162P3} gp120 were determined by enzyme-linked immunosorbent assay (ELISA)¹⁹. Neutralization of monoclonal antibodies and sera were performed either with phytohaemagglutinin-activated peripheral blood mononuclear cells (PBMC) from a single rhesus macaque as target cells¹⁹ or using an HIV-1 envelope pseudotyped luciferase SHIV_{SF162P3} assay by Monogram Biosciences²⁹. C1q binding and C3 fixation using b12 or variants and human or macaque serum were measured in ELISA. Human $Fc\gamma$ RIa, $Fc\gamma$ RIIa (131R), $Fc\gamma$ RIIb and $Fc\gamma$ RIIIa (158V) and macaque $Fc\gamma$ RIa,

FcγRIIa and FcγRIIIa were cloned and expressed as soluble proteins in CHO-K1-SV or HEK293 cells. Binding of antibodies to human and macaque FcγRs was measured in ELISA. Depletion of CD8⁺ lymphocytes was accomplished using the mouse-human chimaeric anti-CD8 monoclonal antibody, cM-T807²³. The ADCVI assays were based on methods previously described²⁴ using target cells, consisting of CEM.NKR-CCR5 cells, infected with SHIV_{SF162P3} and either huPBMCs or rhPBMCs as effectors. Statistical analysis was performed using GraphPad Prism for Windows, version 4.03. Analysis was performed with a repeated measures two-way ANOVA, followed by Bonferroni post hoc test. The alpha level was 0.05.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 April; accepted 20 July 2007.

- McMichael, A. J. HIV vaccines. *Annu. Rev. Immunol.* **24**, 227–255 (2006).
- Johnston, M. I. & Fauci, A. S. An HIV vaccine—evolving concepts. *N. Engl. J. Med.* **356**, 2073–2081 (2007).
- Mascola, J. R. Passive transfer studies to elucidate the role of antibody-mediated protection against HIV-1. *Vaccine* **20**, 1922–1925 (2002).
- Veazey, R. S. *et al.* Prevention of virus transmission to macaque monkeys by a vaginally applied monoclonal antibody to HIV-1 gp120. *Nature Med.* **9**, 343–346 (2003).
- Parren, P. W. & Burton, D. R. The antiviral activity of antibodies *in vitro* and *in vivo*. *Adv. Immunol.* **77**, 195–262 (2001).
- Klasse, P. J. & Sattentau, Q. J. Occupancy and mechanism in antibody-mediated neutralization of animal viruses. *J. Gen. Virol.* **83**, 2091–2108 (2002).
- Burton, D. R. Antibodies, viruses and vaccines. *Nature Rev. Immunol.* **2**, 706–713 (2002).
- Huber, V. C., Lynch, J. M., Bucher, D. J., Le, J. & Metzger, D. W. Fc receptor-mediated phagocytosis makes a significant contribution to clearance of influenza virus infections. *J. Immunol.* **166**, 7381–7388 (2001).
- Baldrige, J. R. & Buchmeier, M. J. Mechanisms of antibody-mediated protection against lymphocytic choriomeningitis virus infection: mother-to-baby transfer of humoral protection. *J. Virol.* **66**, 4252–4257 (1992).
- Gauduin, M. C., Weir, R., Fung, M. S. & Koup, R. A. Involvement of the complement system in antibody-mediated post-exposure protection against human immunodeficiency virus type 1. *AIDS Res. Hum. Retroviruses* **14**, 205–211 (1998).
- Binley, J. M. *et al.* Passive infusion of immune serum into simian immunodeficiency virus-infected rhesus macaques undergoing a rapid disease course has minimal effect on plasma viremia. *Virology* **270**, 237–249 (2000).
- Gomez-Roman, V. R. *et al.* Vaccine-elicited antibodies mediate antibody-dependent cellular cytotoxicity correlated with significantly reduced acute viremia in rhesus macaques challenged with SIVmac251. *J. Immunol.* **174**, 2185–2189 (2005).
- Flores, R. H. *et al.* Evaluation of passively transferred, nonneutralizing antibody-dependent cellular cytotoxicity-mediating IgG in protection of neonatal rhesus macaques against oral SIVmac251 challenge. *J. Immunol.* **177**, 4028–4036 (2006).
- Demberg, T. *et al.* Non-neutralizing antibodies and vaccine-induced protection. *Retrovirology* **3** (Suppl. 1), S26 (2006).
- Holl, V. *et al.* Efficient inhibition of HIV-1 replication in human immature monocyte-derived dendritic cells by purified anti-HIV-1 IgG without induction of maturation. *Blood* **107**, 4466–4474 (2006).
- Holl, V. *et al.* Nonneutralizing antibodies are able to inhibit human immunodeficiency virus type 1 replication in macrophages and immature dendritic cells. *J. Virol.* **80**, 6177–6181 (2006).
- Burton, D. R. *et al.* Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science* **266**, 1024–1027 (1994).
- Binley, J. M. *et al.* Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies. *J. Virol.* **78**, 13232–13252 (2004).
- Parren, P. W. *et al.* Antibody protects macaques against vaginal challenge with a pathogenic R5 simian/human immunodeficiency virus at serum levels giving complete neutralization *in vitro*. *J. Virol.* **75**, 8340–8347 (2001).
- Zhou, T. *et al.* Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445**, 732–737 (2007).
- Hezareh, M., Hessel, A. J., Jensen, R. C., van de Winkel, J. G. & Parren, P. W. Effector function activities of a panel of mutants of a broadly neutralizing antibody against human immunodeficiency virus type 1. *J. Virol.* **75**, 12161–12168 (2001).
- Nimmerjahn, F. & Ravetch, J. V. Fcγ receptors: old friends and new family members. *Immunity* **24**, 19–28 (2006).
- Schmitz, J. E. *et al.* A nonhuman primate model for the selective elimination of CD8⁺ lymphocytes using a mouse-human chimeric monoclonal antibody. *Am. J. Pathol.* **154**, 1923–1932 (1999).
- Forthal, D. N. *et al.* Rhesus macaque polyclonal and monoclonal antibodies inhibit simian immunodeficiency virus in the presence of human or autologous rhesus effector cells. *J. Virol.* **80**, 9217–9225 (2006).
- Harouse, J. M. *et al.* Mucosal transmission and induction of simian AIDS by CCR5-specific simian/human immunodeficiency virus SHIV(SF162P3). *J. Virol.* **75**, 1990–1995 (2001).
- Harouse, J. M., Gettie, A., Tan, R. C., Blanchard, J. & Cheng-Mayer, C. Distinct pathogenic sequela in rhesus macaques infected with CCR5 or CXCR4 utilizing SHIVs. *Science* **284**, 816–819 (1999).
- Tan, R. C., Harouse, J. M., Gettie, A. & Cheng-Mayer, C. *In vivo* adaptation of SHIV(SF162): chimeric virus expressing a NSI, CCR5-specific envelope protein. *J. Med. Primatol.* **28**, 164–168 (1999).
- Marx, P. A. *et al.* Progesterone implants enhance SIV vaginal transmission and early virus load. *Nature Med.* **2**, 1084–1089 (1996).
- Richman, D. D., Wrinn, T., Little, S. J. & Petropoulos, C. J. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl Acad. Sci. USA* **100**, 4144–4149 (2003).
- Zwick, M. B. *et al.* Identification and characterization of a peptide that specifically binds the human, broadly neutralizing anti-human immunodeficiency virus type 1 antibody b12. *J. Virol.* **75**, 6692–6699 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Saye-Francisco, P. and C. Carney, R. Aguilar-Sino and D. Tehrani for antibody production assistance at TSRI. We also thank T. Vink for expressing FcγR and the technical assistance of A. van den Broek and A. Ortiz Buijsse at Genmab. We are grateful for the assistance provided by C. Corbaci during the preparation of the manuscript. We also thank M. Zwick and R. Pantophlet for discussions. Support for this work was provided by an NIH grant (D.R.B.), by the Neutralizing Antibody Consortium of the International AIDS Vaccine Initiative, and by a Swiss National Foundation Fellowship (L.H.) and an NIH grant (D.N.F.).

Author Contributions Project planning was performed by A.J.H., L.H., P.W.H.I.P., P.A.M., D.R.B.; experimental work by A.J.H., L.H., M.H., C.E.G.H., F.J.B., G.L., D.N.F.; data analysis by A.J.H., L.H., J.M.B., C.M.S.L., G.L., D.N.F., P.W.H.I.P., P.A.M., D.R.B.; and manuscript composition by A.J.H., L.H., P.W.H.I.P. and D.R.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.R.B. (burton@scripps.edu).

METHODS

Macaques. All protocols for female Indian rhesus macaques were reviewed and approved by the Institutional Animal Care and Use Committees. The animals were housed in accordance with the American Association for Accreditation of Laboratory Animal Care Standards. At the start of all experiments except the CD8⁺ T-cell depletion studies, all animals were experimentally naive and were negative for antibodies against HIV-1, SIV and type D retrovirus. Virus challenges and intravenous antibody transfers are more fully described elsewhere⁴.

Challenge virus. The virus used in this study was SHIV_{SF162P} passage 3, which has been described elsewhere^{25–27}. SHIV_{SF162P3} retains the R5 phenotype of HIV-1_{SF162}. SHIV_{SF162P3}, propagated in phytohaemagglutinin (PHA)-activated rhesus macaque peripheral blood mononuclear cells (PBMCs), was obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH (catalogue no. 6526; contributors: J. Harouse, C. Cheng-Mayer and R. Pal).

b12 and variant antibodies. IgG1 b12 is a human antibody (IgG1, κ) that recognizes an epitope overlapping the CD4-binding site of gp120 (refs 17, 18). Variants of b12 were created by site-directed mutagenesis as previously described²¹.

Antibody production. Recombinant IgG1 (wild-type b12, isotype control and b12 variants K322A and L234A, L235A) were expressed in Chinese hamster ovary (CHO-K1) cells in glutamine-free custom formulated Glasgow minimum essential medium (GMEM Selection Media) (MediaTech Cellgro). Two IgG1 isotype control antibodies were used in this study: (1) an anti-dengue NS1 antibody, and (2) an anti-ebola GP antibody. For large-scale tissue culture, media was supplemented with 3.5% Ultra Low Bovine IgG Fetal Bovine Serum (Invitrogen) and grown in ten-layer Cellstacks and Cell Cubes (Corning). Antibodies were purified using Protein A affinity matrix (GE Healthcare), and dialysed against phosphate-buffered saline (PBS). Care was taken to minimize endotoxin contamination, which was monitored using a quantitative chromogenic Limulus Amoebocyte Lysate assay (Cambrex) performed according to the manufacturer's recommendations. Antibody used for the passive transfer experiments contained <3 IU of endotoxin per ml.

Plasma viral loads. Quantitative assays for the measurement of SIV RNA were performed at Siemens Medical Solutions (formerly Bayer Diagnostics) using a branched DNA signal amplification assay for SIV²⁸.

Serum antibody ELISAs. Antibody concentrations in macaque sera were determined by two different ELISAs: (1) recombinant monomeric HIV-1 gp120_{JR-FL} (provided by Progenics), and (2) the B2.1 peptide. B2.1 is a homodimer of the peptide HERSYMFSDLENRCI-(biotinylated Orn)-KK (synthesized by AnaSpec). This peptide binds the b12-antigen-binding site with high specificity as described in detail³⁰. The ELISAs are fully described elsewhere¹⁹.

SHIV capture ELISA. Measurement of binding to SHIV_{SF162P3} gp120 was determined in Rhesus (no. 355) PBMC-derived viral culture lysate by capture with the anti-gp120 antibody D7324 (International Enzymes) in ELISA. Microtitre plates (Corning) were coated overnight at 4 °C with D734 at a concentration of 5 $\mu\text{g ml}^{-1}$ diluted in phosphate-buffered saline (PBS). All subsequent steps were performed at room temperature. Unless otherwise specified, plates were washed four times after each step with PBS supplemented with 0.05% Tween. After incubation and washing, coated plates were blocked for 1 h with PBS supplemented with 3% bovine serum albumin (BSA). Blocking solution was decanted and viral culture supernatants diluted 1:3 in PBS containing 1% BSA and 0.02% Tween (PBS-B-T) were added and incubated for 2 h. After washing, antibodies serially diluted in PBS-B-T (5 $\mu\text{g ml}^{-1}$ –0.32 ng ml^{-1}) were added and incubated for 2 h. Plates were again washed and alkaline-phosphatase-conjugated goat anti-human IgG F(ab')₂ (Pierce) was added (diluted 1:500 in PBS-B-T) and incubated for 1 h. Plates were washed and developed with *p*-nitrophenyl phosphate (Sigma) and absorbances were read at 405 nm at 10 min, 20 min and 30 min. The assay volume for each step was 50 μl .

Neutralization assays. Neutralization of antibodies and sera was assessed by two different methods. Neutralization of the primary isolate SHIV_{SF162P3} was performed using phytohaemagglutinin (PHA)-activated PBMCs from a single rhesus macaque (no. 355) as target cells. Cells from this animal replicate SHIV_{SF162P} efficiently. Neutralization assessment was carried out as described previously¹⁹. Neutralization titres of animal sera were reported by Monogram Biosciences after preparation of an HIV-1 envelope pseudotyped luciferase SHIV_{SF162P3} capable of single-round replication. The pseudovirus-based neutralization assay was performed as previously described²⁹.

C1q binding and C3 fixation assays. The ability of b12 and variants to bind C1q or cause fixation of C3 present in human and macaque serum was measured in ELISA. To assess the affinity for C1q of b12 and variants present in human and macaque serum, antibodies were serially diluted (62.5–0.2 $\mu\text{g ml}^{-1}$) and coated onto a microtitre plate and incubated overnight at 4 °C. Plates were blocked with 0.5× PBST/1% gelatin followed by the addition of human or rhesus serum (3%). After washing, rabbit anti-human C1q (DAKO) was added at a dilution of 1:1,000. IgG-bound C1q was detected with a peroxidase-conjugated swine anti-rabbit IgG (DAKO) and was subsequently developed with ABTS. For the C3 fixation assay, mouse anti-human C3 (DAKO) was added at a dilution of 1:100 followed by detection with peroxidase-conjugated rabbit anti-mouse IgG (Jackson) diluted 1:1,000. Reactions were stopped with 2% oxalic acid after 10 min. PBST washes were carried out between each step. The assay volume for each step was 100 μl . Unless otherwise specified, incubation times were 1 h at room temperature with gentle shaking.

Fc γ R binding assays. All Fc γ R proteins contained a histidine tail for purification. Binding of antibodies to human and macaque Fc γ Rs was measured in ELISA by capturing the recombinant Fc γ Rs with an anti-poly-histidine antibody (Research Diagnostics) at 4 $\mu\text{g ml}^{-1}$ diluted in PBS, coated onto a microtitre plate and incubated overnight at 4 °C. Fc γ Rs were added and incubated at a final concentration of 2 $\mu\text{g ml}^{-1}$ in PBST. After washing, serial dilutions of b12, variants and isotype controls were added. HRP-labelled F(ab')₂ fragments of goat anti-human F(ab')₂ fragments of IgG (Jackson) were used as the detection conjugate (incubation for 90 min). The results were visualized with ABTS and the reaction was stopped with 2% oxalic acid. PBST washes were carried out between each step. The assay volume for each step was 100 μl . Unless otherwise specified, incubation times were 75 min at room temperature.

CD8⁺ T-cell depletions. Depletion of CD8⁺ lymphocytes was accomplished using the mouse-human chimaeric anti-CD8 monoclonal antibody, cM-T807 (provided by the NIH Nonhuman Primate Resource, Beth Israel Deaconess Medical Center, Boston, Massachusetts). Administration of the antibody was given on days 0, 3 and 7 at doses of 10 mg kg^{-1} on day 0, followed by doses of 5 mg kg^{-1} on days 3 and 7. Methods for depletion and confirmation of depletion are fully described elsewhere²³.

ADCVI assays. The ADCVI assays were based on methods previously described²⁴. Targets, consisting of CEM.NKR-CCR5 cells, were infected with SHIV_{SF162P3} and washed to remove cell-free virus and incubated for 48 h. (CEM.NKR-CCR5 cells were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH (catalogue no. 4376; contributor: A. Trkola.) Effector cells, either huPBMCs or rhPBMCs, were combined with serially diluted antibodies and added to target cells for a 30-min incubation. Cells were re-washed and incubated for 7 days at 37 °C in 5% CO₂. Supernatant was collected and assayed for p27 by ELISA (Zeptometrix).

Statistics. The control group consisted of 4 animals ($n = 4$), and each of the treated groups consisted of 9 animals ($n = 9$). For practical reasons, the experiment was divided into two parts with exactly the same set up. In experiment 1, there were 2 animals in the control group and 4 animals in each of the treated groups. In experiment 2, there were 2 animals in the control group and 5 in each of the treated groups. Statistical analysis was performed using GraphPad Prism for Windows, version 4.03 (GraphPad Software, 2005). Group differences in viral load in time and group differences in neutralization titres were determined after normalization of data by log₁₀ transformation.

p53 is regulated by the lysine demethylase LSD1

Jing Huang¹, Roopsha Sengupta², Alexandra B. Espejo³, Min Gyu Lee¹, Jean A. Dorsey¹, Mario Richter², Susanne Opravil², Ramin Shiekhattar¹, Mark T. Bedford³, Thomas Jenuwein² & Shelley L. Berger¹

p53, the tumour suppressor and transcriptional activator, is regulated by numerous post-translational modifications, including lysine methylation^{1,2}. Histone lysine methylation has recently been shown to be reversible; however, it is not known whether non-histone proteins are substrates for demethylation. Here we show that, in human cells, the histone lysine-specific demethylase LSD1 (refs 3, 4) interacts with p53 to repress p53-mediated transcriptional activation and to inhibit the role of p53 in promoting apoptosis. We find that, *in vitro*, LSD1 removes both monomethylation (K370me1) and dimethylation (K370me2) at K370, a previously identified Smyd2-dependent monomethylation site². However, *in vivo*, LSD1 shows a strong preference to reverse K370me2, which is performed by a distinct, but unknown, methyltransferase. Our results indicate that K370me2 has a different role in regulating p53 from that of K370me1: K370me1 represses p53 function, whereas K370me2 promotes association with the coactivator 53BP1 (p53-binding protein 1) through tandem Tudor domains in 53BP1. Further, LSD1 represses p53 function through the inhibition of interaction of p53 with 53BP1. These observations show that p53 is dynamically regulated by lysine methylation and demethylation and that the methylation status at a single lysine residue confers distinct regulatory output. Lysine methylation therefore provides similar regulatory complexity for non-histone proteins and for histones.

Previous observations of p53 methylation led us to test whether LSD1, also known as BHC110 (ref. 3), has a role in p53 signalling. We used co-immunoprecipitation to examine whether LSD1 binds to p53. Ectopically expressed tagged LSD1 and p53 associated in HEK-293 cells (Fig. 1a). HDAC2, which is present in a stable complex with LSD1 (ref. 5), also co-immunoprecipitated with p53 (Fig. 1a). Using nuclear extract, we detected interaction between endogenous LSD1 and p53 in MCF7 cells with or without treatment by adriamycin, a DNA-damaging reagent that activates p53 (Fig. 1b). Similar results were obtained in HEK-293 cells (Supplementary Fig. 1).

We studied whether LSD1 interacts directly with p53. Purified recombinant Flag–LSD1 from insect cells (Supplementary Fig. 3a) and glutathione S-transferase (GST)-tagged p53 from bacteria were used in GST pull-down assays. We observed that GST–p53, but not GST alone, associates with Flag–LSD1, suggesting a direct interaction between p53 and LSD1 (Fig. 1c).

LSD1 can act either as a transcriptional coactivator or as a co-repressor^{4,6}. We investigated the role of LSD1 in p53-mediated transcriptional regulation in either mock-treated or adriamycin-treated U2OS cells. Ablation of LSD1 with short interfering RNA (siRNA) results in increased expression of *p21* and *mdm2* (Fig. 2a), two well-characterized p53 target genes. The increased expression of *p21* and *mdm2* was not caused by an increased steady-state level of p53 after treatment with LSD1 siRNA (Supplementary Fig. 2a). In fact, we consistently observed a slight decrease in the steady-state level of p53 without adriamycin treatment, which probably resulted from feedback proteolysis of p53 caused by an elevated level of *mdm2*

when that of LSD1 was decreased. Similar results were observed in MCF7 cells (Supplementary Fig. 2b, c). To examine whether the effect of LSD1 knockdown on the expression of *p21* and *mdm2* is p53-dependent, we used the cell lines BJ and BJ-DNp53 (refs 2, 7). BJ-DNp53 cells contain a stably incorporated dominant-negative p53 (DNp53) that prevents endogenous p53 from binding to DNA^{7,8}. We observed that a decrease in LSD1 level increased the expression of *p21* in BJ cells, but not in BJ-DNp53 cells (Fig. 2b and Supplementary Fig. 2d). A similar dependence of LSD1-mediated repression on p53 was observed in another pair of cell lines, HCT116 (*p53*^{+/+}) and HCT116 (*p53*^{-/-}) (Supplementary Fig. 2e). Taken together, these results show that LSD1 represses the transcriptional activity of p53.

Because LSD1 does not regulate the activity of p53 in BJ-DNp53 cells in which the DNA-binding ability of p53 is impaired, we reasoned that LSD1 might affect the binding of DNA by p53, thus

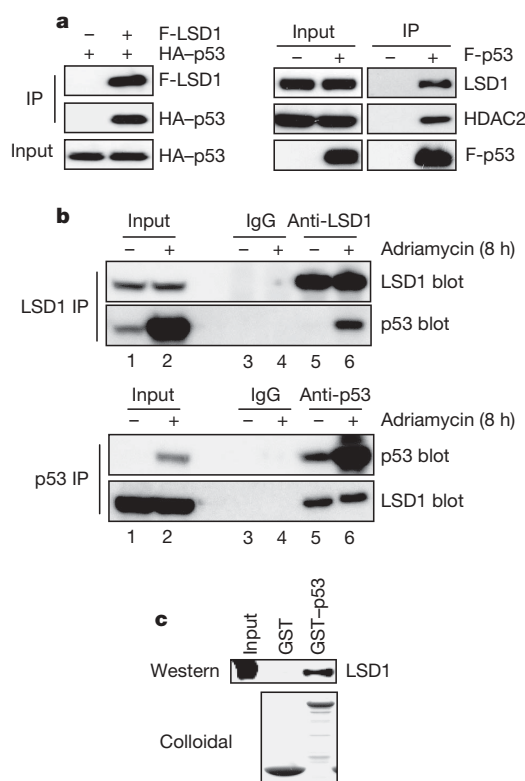


Figure 1 | LSD1 interacts with p53. **a**, **b**, Reciprocal co-immunoprecipitation assays were used to detect interaction between ectopic (**a**) and endogenous (**b**) LSD1 and p53 in HEK-293 cells (**a**) and MCF7 cells (**b**). **c**, GST pull-down assay. F, Flag; HA, haemagglutinin; IP, immunoprecipitation.

¹The Wistar Institute, 3601 Spruce Street, Philadelphia, Pennsylvania 19104, USA. ²Research Institute of Molecular Pathology (IMP), The Vienna Biocenter, 1030 Vienna, Austria. ³M. D. Anderson Cancer Center, Department of Carcinogenesis, University of Texas, Smithville, Texas 78957, USA.

altering the equilibrium between DNA-bound and free p53. To test this hypothesis, we used chromatin immunoprecipitation (ChIP) and real-time polymerase chain reaction (PCR). We found significantly higher levels of p53 at the *p21* promoter in U2OS cells than in control cells when the level of LSD1 was decreased by siRNA treatment (Fig. 2c). These results indicate that LSD1 represses the activity of p53 in part through decreasing the binding of p53 to DNA.

After DNA damage, p53 activates the expression of genes that are involved in either cell cycle arrest or apoptosis. Because LSD1 represses the transcriptional activation activity of p53, we proposed that LSD1 inhibits p53-mediated apoptosis and/or cell cycle arrest. Without adriamycin treatment, decreasing LSD1 levels with lentivirus-based short hairpin RNA (shRNA) did not affect the percentage of apoptotic cells (sub-G1 population) in U2OS cells (Fig. 2d, upper panels, and Supplementary Fig. 2f). However, the apoptotic cell population increased (to 19%) in response to adriamycin treatment, and increased further (to 33.3%) when LSD1 levels were decreased with shRNA (Fig. 2d, lower panels). These results indicate that LSD1 represses p53-mediated apoptosis.

LSD1 was the first lysine demethylase to be characterized; it demethylates lysine 4 or lysine 9 on histone H3 (refs 4–6). Because LSD1 binds directly to p53 (Fig. 1c), we examined whether LSD1 demethylates p53. To assay the two known p53 methylation sites *in vitro*, we developed a methylation–demethylation assay (Supplementary Fig. 3a–c and Methods). We found that LSD1 specifically removes Smyd2-mediated methylation but not methylation catalysed by Set9 (Fig. 3a). Because K370 is the only amino acid residue modified by Smyd2 *in vitro* (Supplementary Fig. 3d), we conclude that LSD1 demethylates p53 at K370 *in vitro*.

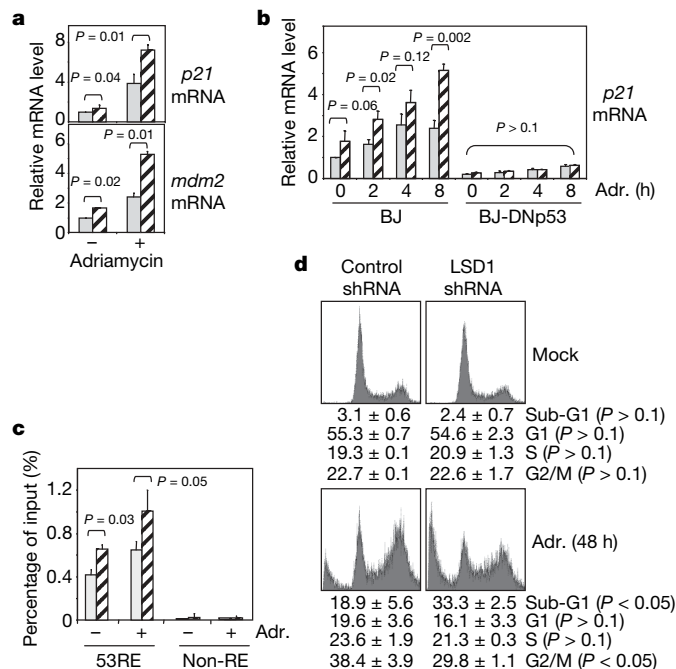


Figure 2 | LSD1 represses the activity of p53. **a**, Quantitative real-time PCR to assess the relative mRNA levels of *p21* and *mdm2* in U2OS cells transfected with control (grey columns) or LSD1 (hatched columns) siRNA followed by 24 h of mock treatment or treatment with adriamycin. **b**, Quantitative real-time PCR was used to measure the relative mRNA levels of *p21* in BJ and BJ-DNP53 cells transfected with control (grey columns) or LSD1 (hatched columns) siRNA followed by treatment with adriamycin (Adr.) for 0, 2, 4 or 8 h. **c**, ChIP assay to measure p53 bound to its response element (53RE) and to a control site (non-RE) within the *p21* promoter in U2OS cells transfected with control (grey columns) or LSD1 (hatched columns) siRNA followed by treatment with adriamycin for 8 h. **d**, Flow cytometry analysis to assess the effect of LSD1 on the percentage of apoptotic (that is, sub-G1) U2OS cells. Results are shown as means \pm s.d.

Smyd2 is a monomethyl transferase², and LSD1 is able to remove both monomethylation and dimethylation from histone substrates⁴. To determine whether LSD1 can remove dimethylation at p53K370, we used peptides representing monomethylated (K370me1), dimethylated (K370me2) and trimethylated (K370me3) substrate, as well as monomethylated (K372me1) p53K372 (Fig. 3b). Reaction products were subjected to dot-blot analysis to assess the different methylation sites and levels of methylation by using specific antibodies^{1,2}. We found that LSD1 decreases the level of K370me1 and K370me2 but not that of K370me3 (Fig. 3b), supporting previous findings that LSD1 demethylates only monomethylated and dimethylated lysine residues on histone substrates⁴. Consistent with the observation on methylated GST–p53 (Fig. 3a) was our observation that LSD1 was not able to demethylate K372me1 peptide (Fig. 3b). These results show that LSD1 specifically removes both K370me1 and K370me2 *in vitro*.

We next investigated whether LSD1 affects K370 or K372 methylation *in vivo*. The level of LSD1 was decreased in U2OS cells by using siRNA, and then K370me1, K370me2 and K372me1 signals were determined by western blot analysis after immunoprecipitation with p53 antibody. On adriamycin treatment, a decrease in LSD1 level significantly increased that of K370me2, whereas that of K370me1

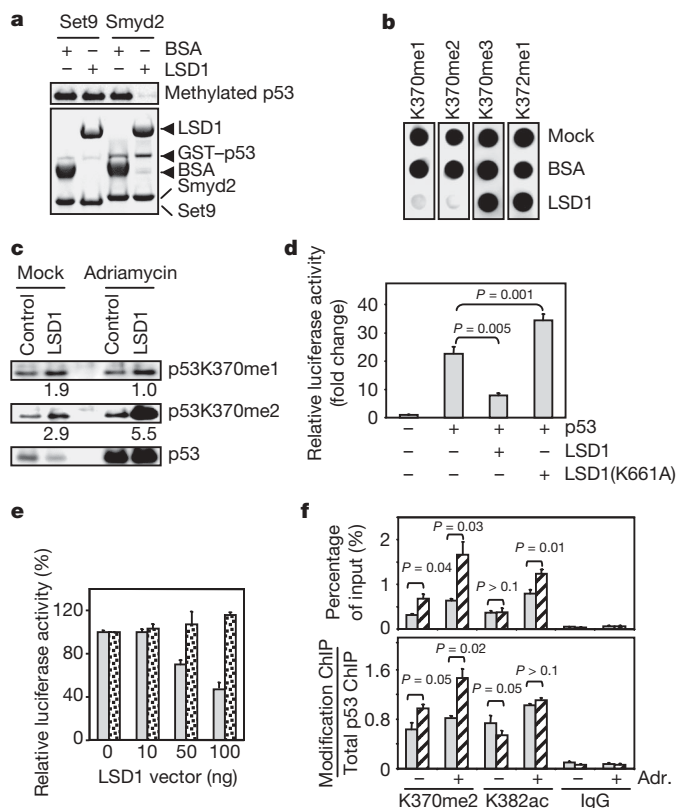


Figure 3 | LSD1 demethylates p53 at K370. **a**, Top: fluorography, showing demethylation of p53 by LSD1. Bottom: colloidal staining, showing loadings. **b**, Dot-blot analysis, showing demethylation of p53 peptides by LSD1. **c**, Western blot analysis of DO1 immunoprecipitation in U2OS cells transfected with control or LSD1 siRNA followed by treatment with adriamycin for 8 h. Numbers below the western blots show the ratio of the signal with LSD1 knockdown to control knockdown, then to p53 levels. **d**, **e**, Dual luciferase assay in H1299 cells with a reporter bearing *p21* promoter to show that LSD1-mediated repression is dependent on its demethylation activity (**d**) and K370 in p53 (**e**). In **e**: grey columns, wild-type p53; stippled columns, p53(K370R). **f**, ChIP assay to detect p53K370me2 and p53K382ac recruitment to *p21* promoter in U2OS cells transfected with control (grey columns) or LSD1 (hatched columns) siRNA followed by treatment with adriamycin (Adr.) for 8 h. Top: absolute percentage of input. Bottom: ratio of modification ChIP to total p53 ChIP. Results in **d–f** are shown as means \pm s.d.

was affected only slightly (Fig. 3c). Therefore, although LSD1 is able to remove both K370me1 and K370me2 *in vitro*, it seems to demethylate K370me2 preferentially *in vivo*. Other histone demethylases also exhibit a preference for specific methylation levels^{9,10}. The underlying mechanism for these preferences is unclear but may involve other post-translational modifications on the substrate or demethylase-associated proteins. Under these assay conditions, endogenous K370me3 and K372me1 signals are not detected, even after long exposures in western blot analysis (data not shown). We note that Smyd2 is a monomethyl transferase²; our results therefore indicate that K370me2 is performed by a distinct, currently unidentified methyltransferase. Taken together, our results reveal that p53K370me2 is demethylated by LSD1.

To assess whether the demethylation activity of LSD1 is required for its repression of the transcriptional activity of p53, we ectopically expressed *LSD1* and *LSD1*(K661A), a demethylation-defective mutant⁵, together with p53 in H1299 cells. A luciferase reporter bearing the *p21* promoter was used to assess transcriptional activation driven by p53 (Fig. 3d). We observed that LSD1 but not LSD1(K661A) decreases the activity of p53 (Fig. 3d). LSD1 and LSD1(K661A) interact with p53 comparably (Supplementary Fig. 3e), suggesting that LSD1 represses the activity of p53 in a demethylation-dependent manner. To test whether LSD1 represses the activity of p53 in a K370-dependent manner, we co-transfected H1299 cells with vectors expressing *LSD1* along with either p53 or p53(K370R). We found that LSD1 decreased luciferase activity mediated by wild-type p53 but not by p53(K370R) (Fig. 3e), demonstrating that LSD1 represses the activity of p53 through K370.

We determined whether p53 dimethylated at K370 is present at the *p21* promoter in U2OS cells by ChIP assay with DO1 (unmodified p53), p53K370me2 and p53K382 acetylation (K382ac) antibodies (Fig. 3f). The K382ac antibody serves as a positive control because it detects a well-characterized activating modification of p53 during DNA damage^{11,12}. Treatment with adriamycin results in increased K370me2 and K382ac ChIP signals (Fig. 3f, grey bars), suggesting that K370me2 is an activating modification for p53. Decreasing the level of LSD1 by siRNA results in a strongly increased K370me2 signal, whereas the K382ac signal is enhanced only slightly (Fig. 3f, upper panel). The K370me2 level increases even after normalization to the total p53 level (Fig. 3f, lower panel), showing that higher K370me2 is due not only to elevated p53 levels at the promoter caused by treatment with LSD1 siRNA (Fig. 2c). Thus, p53K370me2 increases at the *p21* promoter after DNA damage and this recruitment is regulated by LSD1.

These results suggest that K370me2 may have a distinct role in regulating p53 in comparison with repression-linked Smyd2-mediated K370me1. Recent studies show that specific lysine methylation states in histones serve as recognition sites for specific binding proteins^{13–17}. We investigated binding proteins that recognize specific methylation states at K370 by screening a GST protein domain microarray¹⁸ (Fig. 4a and Supplementary Fig. 4). We found that the tandem Tudor domains of 53BP1 preferentially recognize K370me2 peptide, compared with K370me0, K370me1 or K370me3 (Fig. 4a). We used peptide pull-down to confirm that 53BP1 binds to the K370me2 peptide more strongly than to K370me1 or K370me3 peptide, and that there is no binding to the K370me0 peptide (Supplementary Fig. 5a). These results are similar to recent observations *in vitro* that the tandem Tudor domains of 53BP1 preferentially bind to histone dimethylated at H3K79 and H4K20, and bind more weakly to monomethylated H3K79 and H4K20 (refs 18–20).

We next examined whether 53BP1 associates with dimethylated p53K370 more strongly than with monomethylated p53K370 *in vivo*. We performed a Flag immunoprecipitation assay to pull down Flag–53BP1-associated proteins followed by western blot analysis to detect p53K370me1 and K370me2 (Fig. 4b). Because the p53K370me1 signal is weaker than that of K370me2, *Smyd2* was coexpressed with p53 to provide equivalent K370me1 input. Under the same loading

conditions of input signals, we detected more K370me2 signal than K370me1 in the Flag–53BP1 immunoprecipitation eluates (Fig. 4b). These results *in vivo* support the hypothesis that 53BP1 binds to p53K370me2 more strongly than to p53K370me1.

Physical interaction between p53 and 53BP1 was discovered in a yeast two-hybrid analysis²¹, and *in vivo* interaction between p53 and 53BP1 has recently been reported^{22,23}. We are able reproducibly to detect interaction between endogenous p53 and 53BP1 *in vivo* (Supplementary Fig. 5b). The specificity of this interaction

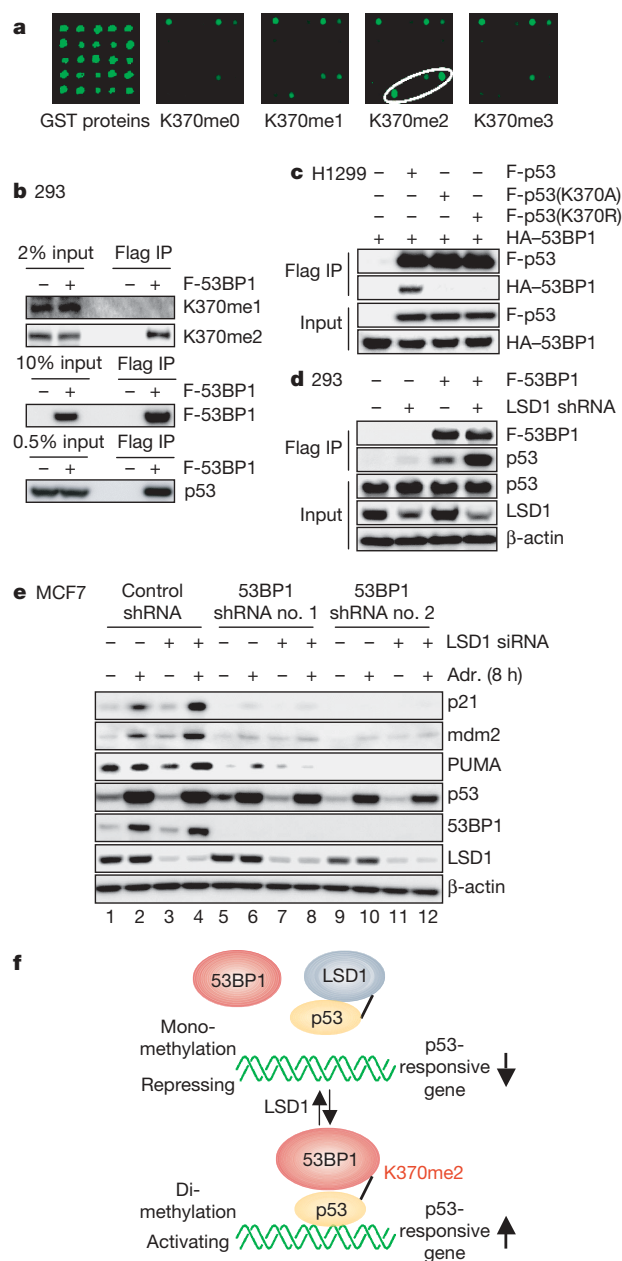


Figure 4 | LSD1 represses the activity of p53 through 53BP1. **a**, GST protein domain array to identify binding proteins for p53K370me0, p53K370me1, p53K370me2 and p53K370me3. The K370me2-dependent interaction with tandem Tudor domains of 53BP1 is circled. Part of the protein array (C3 of Tudor domain section) is shown. For the entire array, see Supplementary Fig. 4. **b**, Flag immunoprecipitation (IP) to assay the relative binding of 53BP1 to K370me1 and K370me2 *in vivo*. **c**, Flag IP to determine the requirement of K370 for the interaction between 53BP1 and p53. **d**, Flag IP to study the effect of LSD1 knockdown on the 53BP1–p53 interaction. **e**, Western blot analysis to test whether LSD1-mediated repression is 53BP1-dependent. Adr., adriamycin. **f**, Schematic model for the dynamic regulation of K370 methylation.

was analysed in several approaches. First, we generated 53BP1 bearing substitution mutations in conserved residues in the Tudor domains previously shown to be required for binding to histone H3K79me2 and H4K20me2 (Y1502Q or D1521R)^{19,20}. These substitutions in 53BP1 strongly decreased co-immunoprecipitation with p53 (Supplementary Fig. 5c). Second, p53 substituted at K370 (K370A or K370R) showed greatly decreased interaction with 53BP1 (Fig. 4c). Last, decreasing the level of LSD1 with shRNA increased the interaction between p53 and 53BP1 (Fig. 4d), indicating that LSD1 represses the activity of p53 through decreasing the interaction with 53BP1.

53BP1 has been functionally linked to p53 as a potential coactivator^{21,24,25}. We directly assayed whether 53BP1 is required for transcriptional activation by p53. 53BP1 levels in MCF7 cells were stably reduced by lentivirus-based shRNAs (Fig. 4e). Two different shRNAs targeting 53BP1 for knockdown resulted in decreased protein levels of p21, mdm2 and PUMA (p53 upregulated modulator of apoptosis) (Fig. 4e, compare lane 2 with lane 6 or 10) and messenger RNA (Supplementary Fig. 5d). As described above (Fig. 2 and Supplementary Fig. 2), treatment with LSD1 siRNA increased the expression of p21, mdm2 and PUMA in MCF7 cells (Fig. 4e, compare lanes 2 and 4). In contrast, gene expression did not increase in the LSD1 knockdown in cells also carrying 53BP1 shRNAs (Fig. 4e, compare lanes 2 and 4 with lanes 6 and 8 or 10 and 12). Together, these results strongly indicate that LSD1 regulates the activity of p53 through its coactivator 53BP1.

Our results demonstrate that LSD1 demethylates p53 at K370 and displays a strong preference *in vivo* for demethylating p53K370me2. In addition, although we did not detect K372me1 as a substrate, previous observation that K372me1 is transient during p53 activation¹ indicates that other demethylases might target methylated p53.

It is well established in histones that various methylation levels elicit specific biological outcomes, through recognition by different binding proteins^{13–15}. In this study we found that K370me2, in contrast to K370me1, has an activating role in p53 regulation through providing an interaction surface for the binding of 53BP1. These results add another layer of complexity to the methylation-mediated regulation of p53. Methylation, similarly to other post-translational modifications of p53, seems to fine-tune p53 function²⁶.

We propose a dynamic model for the regulation of p53 through lysine methylation (Fig. 4f). During gene activation, p53, by means of K370me2, binds to 53BP1. In contrast, during gene repression, LSD1 prevents the accumulation of K370me2 by demethylating the site, thereby disallowing the binding of 53BP1 to p53. The demethylation activity of LSD1 thus maintains p53 in an inactive state, to prevent binding to DNA. 53BP1 and p53 synergize to inhibit tumorigenesis²⁷; our results suggest a mechanism contributing to their cooperative suppression of cancer.

METHODS SUMMARY

Co-immunoprecipitation and GST pull-down assays were used to detect protein–protein interaction. Transient transfection of siRNA and lentivirus transduction of shRNA were used to decrease the levels of specific proteins. Flow cytometry was used to assess the apoptosis and cell cycle arrest caused by p53. ChIP assay was performed to detect the recruitment of p53 and modified p53 to DNA. We used real-time PCR to measure the mRNA level and immunoprecipitated DNA in a ChIP assay. An *in vitro* demethylation assay was used to determine whether LSD1 demethylates p53. A protein domain microarray was used to identify proteins that bind to p53K370me1, p53K370me2 or p53K370me3.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 April; accepted 9 July 2007.

1. Chuikov, S. *et al.* Regulation of p53 activity through lysine methylation. *Nature* **432**, 353–360 (2004).

2. Huang, J. *et al.* Repression of p53 activity by Smyd2-mediated methylation. *Nature* **444**, 629–632 (2006).
3. Hakimi, M. A., Dong, Y., Lane, W. S., Speicher, D. W. & Shiekhattar, R. A candidate X-linked mental retardation gene is a component of a new family of histone deacetylase-containing complexes. *J. Biol. Chem.* **278**, 7234–7239 (2003).
4. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
5. Lee, M. G., Wynder, C., Cooch, N. & Shiekhattar, R. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* **437**, 432–435 (2005).
6. Metzger, E. *et al.* LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436–439 (2005).
7. Vaziri, H. & Benchimol, S. Reconstitution of telomerase activity in normal human cells leads to elongation of telomeres and extended replicative life span. *Curr. Biol.* **8**, 279–282 (1998).
8. Shaulian, E., Zauberman, A., Ginsberg, D. & Oren, M. Identification of a minimal transforming domain of p53: negative dominance through abrogation of sequence-specific DNA binding. *Mol. Cell. Biol.* **12**, 5581–5592 (1992).
9. Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811–816 (2006).
10. Yamane, K. *et al.* JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125**, 483–495 (2006).
11. Barlev, N. A. *et al.* Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases. *Mol. Cell* **8**, 1243–1254 (2001).
12. Gu, W. & Roeder, R. G. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **90**, 595–606 (1997).
13. Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859–872 (2005).
14. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86–90 (2006).
15. Shi, X. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**, 96–99 (2006).
16. Pena, P. V. *et al.* Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**, 100–103 (2006).
17. Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91–95 (2006).
18. Kim, J. *et al.* Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep.* **7**, 397–403 (2006).
19. Huyen, Y. *et al.* Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**, 406–411 (2004).
20. Botuyan, M. V. *et al.* Structural basis for the methylation state-specific recognition of histone H4–K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).
21. Iwabuchi, K., Bartel, P. L., Li, B., Marraccino, R. & Fields, S. Two cellular proteins that bind to wild-type but not mutant p53. *Proc. Natl Acad. Sci. USA* **91**, 6098–6102 (1994).
22. Sengupta, S. *et al.* Functional interaction between BLM helicase and 53BP1 in a Chk1-mediated pathway during S-phase arrest. *J. Cell Biol.* **166**, 801–813 (2004).
23. Ward, I. *et al.* The tandem BRCT domain of 53BP1 is not required for its repair function. *J. Biol. Chem.* **281**, 38472–38477 (2006).
24. Iwabuchi, K. *et al.* Stimulation of p53-mediated transcriptional activation by the p53-binding proteins, 53BP1 and 53BP2. *J. Biol. Chem.* **273**, 26061–26068 (1998).
25. Brummelkamp, T. R. *et al.* An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors. *Nature Chem. Biol.* **2**, 202–206 (2006).
26. Toledo, F. & Wahl, G. M. Regulating the p53 pathway: *in vitro* hypotheses, *in vivo* veritas. *Nature Rev. Cancer* **6**, 909–923 (2006).
27. Morales, J. C. *et al.* 53BP1 and p53 synergize to suppress genomic instability and lymphomagenesis. *Proc. Natl Acad. Sci. USA* **103**, 3310–3315 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Barlev for the Set9 expression vector; D. Reinberg for the p53K372me1 antibody; S. Benchimol for the BJ and BJ-DNP53 cell lines; T. Halazonetis for 53BP1 plasmid; R. Schule for p21 luciferase vector; and members of the T.J. and S.L.B. laboratories for discussions. This project is funded, in part, by a AACR-Pennsylvania Department of Health Fellows grant and Leukemia and Lymphoma Society Special Fellow grant (J.H.). M.T.B. is supported by a Welch Foundation grant. Research in the laboratory of T.J. is supported by the IMP through Boehringer Ingelheim and by grants from the European Union and the Austrian GEN-AU initiative, which is financed by the Austrian Ministry of Education, Science and Culture. Research support to S.L.B. was provided by a grant from the National Cancer Institute at NIH and the Commonwealth Universal Research Enhancement Program of the Pennsylvania Department of Health.

Author Contributions J.H., R.S., A.B.E., M.G.L., J.A.D., M.R. and S.O. performed the experimental work; R.S., M.T.B., T.J. and S.L.B. were responsible for project planning and data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.L.B. (berger@wistar.org).

METHODS

Antibodies. Anti-Smyd2 (amino terminus) and anti-p53K370me1, me2 and me3 antibodies were generated and described previously². Anti-p53K370me2 antibody was antigen-purified from crude serum. Other antibodies were as follows: anti-p53K372me1 (gift from D. Reinberg), anti-LSD1 (Bethyl), anti-p53 (DO1; Santa Cruz), anti-p53 (FL393; Santa Cruz), anti-haemagglutinin (Roche), anti-GST (Upstate), anti-53BP1 (monoclonal; Upstate), anti-53BP1 (polyclonal; Bethyl), anti- β -actin (Sigma), anti-Puma N terminus (Sigma), anti-Flag (Sigma) and mouse and rabbit IgG (Santa Cruz).

Co-immunoprecipitation assay. In the standard co-immunoprecipitation assay, cells were lysed in NET 0.1% buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM EDTA, 0.1% Nonidet P40, freshly added 1 mM phenylmethylsulphonyl fluoride, protease inhibitors) and sonicated with a Bioruptor (Diagenode) for 5 min, with 30-s 'on' and 1-min 'off' cycles. After removal of cell debris by centrifugation, 2 μ g of antibody was added to 1 mg of clarified whole cell extract (WCE) and incubated overnight at 4 °C. The next day, 40 μ l of Protein A-agarose beads (Upstate) was added and incubated for a further 2 h. Beads then were subjected to three washes with NET 0.1% buffer and boiled with 1 \times SDS loading buffer. A modified co-immunoprecipitation assay was used to detect interaction between p53 and 53BP1, in which nuclei were prepared in accordance with the protocol from a Nuclear Extraction Kit (catalogue no. AY2002; Panomics). Nuclear extract (NE) was prepared by lysing the nuclei with NET 0.1% supplemented with 2 mM sodium orthovanadate, 50 mM NaF and 50 mM β -glycerolphosphate as described above for WCE. Typically, 1 mg of NE (3–5 μ g μ l⁻¹) was used for each immunoprecipitation with 2 μ g of specific antibody. For Flag immunoprecipitation, to assess the binding of Flag-53BP1 to K370me1 and K370me2 *in vivo*, *Smyd2* and *p53* were coexpressed with or without Flag-53BP1. The same amounts of input and Flag immunoprecipitation eluates were loaded in western blot analysis for the detection of K370me1 and K370me2.

GST pull-down assay. Bacterial lysate for GST-p53 or GST was pre-bound to 10 μ l of GST beads (Pharmacia) at 25 °C. Beads were then preblocked with *in vitro* pull-down buffer (20 mM HEPES pH 7.9, 150 mM KCl, 1 mM EDTA, 0.1% Nonidet P40, 10% glycerol, 1 μ g μ l⁻¹ BSA and freshly added 1 mM dithiothreitol and 1 mM phenylmethylsulphonyl fluoride) at 4 °C for 1 h. Recombinant Flag-LSD1 (2 μ g) purified from baculovirus-infected *Spodoptera frugiperda* (Sf9) cells was added and incubated for a further 3 h. Beads were washed three times with pull-down buffer, then boiled with 20 μ l of 1 \times SDS loading buffer. Anti-Flag antibody was used to detect Flag-LSD1.

RNA interference. siRNAs were purchased from Dharmacon, and 100 nM siRNA was transfected into cells with DharmaFECT1. siRNA sequences for luciferase control were as described previously². LSD1 siRNA sequences were: 5'-UGAAUUAGCUGAAACACAAUU-3' (sense sequence) and 5'-pUUGUUUUCAGCUAAUUCUU-3' (antisense sequence). Lentivirus-based shRNA vectors for luciferase control and LSD1 were generated with the BLOCK-it U6 RNAi Expression System (Invitrogen). shRNA vectors for 53BP1#1 and 53BP1#2 were purchased from Sigma. The sequences of the short hairpins were: luciferase control, 5'-CACCTAAGGCTATGAAGAGATACCGAAGTATCTCTTCATAGCCTTA-3' (top strand) and 5'-AAAATAAGGCTATGAAGAGATACCTTCGGTATCTCTTCATAGCCTTA-3' (bottom strand); LSD1, 5'-CACCGCACCTTAACAGTGATACTCGAAAGTATCACTGTTATAAGGTGC-3' (top strand) and 5'-AAAAGCACCTTATAACAGTGATACTTCGAGTATCACTGTTATAAGGTGC-3' (bottom strand); 53BP1#1, 5'-CCGGGATACTTGGTCTTACTGTTTCTCGAGAAACCAGTAAGACCAAGTATCTTTTT-3' (top strand)

and 5'-AAAAAGATACTTGGTCTTACTGGTTTCTCGAGAAACCAGTAAGACCAAGTATCCCGG-3' (bottom strand); 53BP1#2, 5'-CCGGCCAGTGTGATTAGTATTGATTCTCGAGAATCAATCAATCACTGGCCCG-3' (bottom strand).

We found that lentivirus-based knockdown of LSD1 by shRNA was not stable and the level of LSD1 recovered two weeks after transduction (data not shown). To overcome this technical difficulty, we repeated the cell transduction two weeks after the first transduction. By this 'double transduction' procedure, we were able to maintain a low level of LSD1 in U2OS, BJ and BJ-DNp53 cells for a month. All the experiments were performed within a month of the first transduction.

Chromatin immunoprecipitation assay. In brief, cells were fixed with 1% formaldehyde and lysed with lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, 0.1% sodium deoxycholate, with protease inhibitors). The cell lysate was sonicated with a Bioruptor machine for a total of 20 min with 30-s 'on' and 1-min 'off' cycles, to shear the DNA to a final size of 200–500 base pairs. After preclearing with Protein A or G beads (Upstate), antibody was added and incubated overnight at 4 °C. The next day, Protein A-agarose or Protein G-agarose beads were added and incubated for a further 1–2 h. The complex was washed twice with lysis buffer, once with high salt buffer (50 mM HEPES-KOH pH 7.5, 500 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, 0.1% sodium deoxycholate), twice with LiCl buffer (10 mM Tris-HCl pH 8.0, 0.25 M LiCl, 0.5% Nonidet P40, 0.5% sodium deoxycholate, 1 mM EDTA) and once with TE buffer, followed by elution in TE buffer containing 1% SDS. The crosslinks were reversed, the DNA was purified with a QIAquick 8 PCR purification kit (Qiagen) and subjected to analysis by quantitative real-time PCR.

In vitro demethylation assay. Recombinant Set9, GST-p53 and LSD1 were purified as described previously^{1,5,11}. Recombinant Smyd2 was purified from insect (Sf9) cells infected with baculovirus. Methylation assays were performed as described previously². GST-p53 (10 μ g) and Set9 (5 μ g) or Smyd2 (5 μ g) were used in a 50- μ l methylation reaction. After passage of the methylation reaction products through a Microspin column 30 (Bio-Rad), 1 μ l of flowthrough was subjected to scintillation counting to measure the level of methylated GST-p53. Comparable scintillation counts within the flowthroughs were used in the demethylation assays with 3–5 μ g of LSD1 in demethylation buffer (50 mM HEPES/NaOH pH 8.0, 25% glycerol)^{4,5,28}. For peptide demethylation, 100 pmol of peptide and 5 μ g of LSD1 purified from bacteria were used in a 50 μ l total volume. After reaction, products were subjected to dot-blot analysis with specific antibodies.

Dual luciferase assay. Dual luciferase assays were performed in accordance with the manufacturer's protocol (Promega). In brief, H1299 cells were seeded in a 24-well plate. At 16 h after transfection, cells were lysed with passive lysis buffer and lysates were subjected to dual luciferase analysis with a Wallac 1420 multi-label counter (PerkinElmer). A *Renilla* luciferase vector (pRL-CMV) was used as an internal control.

Protein domain microarray. Chromo, PhD, Tudor, SANT, MBT and PWWP domains from about 100 chromatin-associated proteins were fused to GST and spotted on a slide¹⁸. Biotin-labelled p53K370me0 (unmethylated), K370me1, K370me2 and K370me3 peptides were individually hybridized to the protein domain array; binding was then detected with Cy5-Streptavidin¹⁸.

28. Forneris, F., Binda, C., Vanoni, M. A., Battaglioli, E. & Mattevi, A. Human histone demethylase LSD1 reads the histone code. *J. Biol. Chem.* **280**, 41360–41365 (2005).

naturejobs

**JOBS OF
THE WEEK**

In the aftermath of the terrorist attacks of 11 September 2001, the United States tightened its visa restrictions. This heightened security led to a fall in the number of overseas graduates applying and being admitted to US colleges and universities. Indeed, a survey by the US Council of Graduate Schools (CGS) showed that between 2003 and 2004 there was a 28% decline in international applications, and a fall of 18% in the number of admission offers. Many feared that a loss of international talent — often seen as the lifeblood of US graduate programmes — would have an adverse effect on educational institutions and jobs.

Many of the visa delays have been addressed as the new procedures have become streamlined, and the number of overseas students applying has begun to rise. But did visa issues alone stymie applications from abroad? Or was interest already waning with international competition on the rise? Perhaps, these factors worked in tandem, with visa problems accelerating the leakage of talent.

The most recent numbers from the CGS, released on 27 August, hint at answers. Offers of admission to prospective international students by US graduate schools rose by 8% from 2006 to 2007 across all fields. This is the third consecutive year of growth — although it is down from last year's 12% increase. Applications in the life sciences increased 18% and offers to applicants grew 11%; in the physical sciences applications grew 12%, offers 8%. To some extent, it seems, universities have recovered — visa restrictions did indeed hinder the flow of talent. But just as interesting was the CGS's documentation of joint and dual degree programmes between US and international universities — a sizeable 30% of graduate schools have them. Competition seems to be globalizing graduate education.

For now, the US graduate-education system continues to be a major draw. But the CGS results neglect to quantify the fact that international students are likely to head home after getting their degrees. After all, there's more international competition not only for prospective students, but also for master's and PhD graduates looking for jobs.

Gene Russo, acting editor of *Naturejobs*

CONTACTS

Acting Editor: Gene Russo

European Head Office, London

The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:

Andy Douglas (4975)
e-mail: a.douglas@nature.com
**Business Development
Manager:**
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com

Natureevents:

Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:

Muriel Lestringuez (4994)

Southwest UK/RoW:

Nils Moeller (4953)

Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

North Germany/The Netherlands:

Reya Silao (4970)

South Germany/Austria:

Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:

Stephen Russell
To send materials use London
address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Jasmine Myer

US Head Office, New York

75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com



Hydrogen hopes

Europe has started to invest in hydrogen, potentially paving the way for a fertile jobs market, says **Quirin Schiermeier**.

Jeremy Rifkin's book *The Hydrogen Economy: The Creation of the Worldwide Energy Web and the Redistribution of Power on Earth* (Tarcher, 2003) paints a strikingly optimistic picture of an energy-efficient world, free of fossil fuels. He foresees power being generated from renewable energies and stored, distributed and shared in the form of hydrogen. This will democratize energy use in a post-fossil-fuel era, says Rifkin, who is president of the Foundation on Economic Trends in Washington DC and an economic adviser to the European Commission. He also expects that the 'hydrogen economy' will create millions of new jobs worldwide in science, engineering, manufacturing, marketing and sales.

Although the jobs machine has a way to go before reaching full speed, multimillion-dollar hydrogen initiatives launched in the United States, Japan and the European Union (EU) suggest a promising future. The EU is a more recent player, but a spike in interest and projected research investments has begun to create new career opportunities in academia and industry for chemists, physicists and engineers.

Much work remains at both the research and the design ends, to take the technology through the lab and demonstration phases and into reliable vehicles and other uses, says John Loughhead, executive director of the UK Energy Research Centre and co-chair of the implementation panel for the EU's 'technology

platform' for hydrogen and fuel cells. "From materials research to advanced mechanical and electrical engineering, there's quite a large demand for experts," Loughhead says. "Unfortunately, there is not a large body of people who combine the various skills required."

The technology's drawbacks have created research challenges. Right now, 'dirty' hydrogen is a problem, as 95% of the gas is currently produced from fossil fuels. Electricity is typically generated by combining hydrogen — taken from fuels such as natural gas, gasoline and diesel — and oxygen, taken from the air, in an electrochemical reaction. Rifkin and others envisage hydrogen being produced from renewable energies such as biomass, an ongoing research area. Others are looking at producing hydrogen via electrolysis on a large scale.

Pushing the limits

The electrode design, catalysts and electrolyte materials used in fuel cells all pose technical problems that limit power and longevity. And much work will also be required for the widespread deployment of fuel-cell technologies, such as refuelling stations.

However, pilot projects are under way. Scandinavian initiatives aim to have consumer hydrogen-fuelled vehicles on roads by 2012. The Scandinavian Hydrogen Highway Partnership was formed in June by Sweden, Denmark and Norway, to establish a network of refuelling stations throughout southern Scandinavia.

The partnership is just one of several demonstration projects aimed at preparing the market for hydrogen as an energy carrier in Europe. Right now only about 20 prototype hydrogen cars a day are being filled with liquid hydrogen at Europe's busiest refuelling station in Lohhof near Munich, Germany.

"We first need reliable figures from the car industry," says Joachim Wolf, executive director of hydrogen



John Loughhead sees a demand for experts.

NREL

solutions for Munich-based Linde Gas and Engineering, which runs the Lohhof station. "We can't put up a fully fledged hydrogen infrastructure and then than wait for the cars to arrive." Pressure to develop affordable hydrogen cars and infrastructure comes mainly from California, pushed by anti-emission laws. "Everybody's targeting the Californian market," says Wolf (see 'Fuelling a new gold rush').

Hydrogen and fuel-cell technologies are based on established disciplines such as electrochemistry, materials science and chemical engineering. Specialist training is usually only provided beyond graduate level. In Britain, for example, master's-level courses are offered at Imperial College London, the University of Oxford and the University of St Andrews. Many other universities, from Taiwan to California, offer a wide range of mainly postgraduate fuel-cell and hydrogen-specific courses (see www.fuelcells.org/ced/career/university.htm).

"All the world is pouncing on hydrogen technology," says Ferdinand Panik, a former head of fuel-cell research at DaimlerChrysler in Stuttgart, Germany. But scientists and engineers are tackling development needs from different perspectives, which can cause problems, he says. "Chemists often have no clue of what's required at the engineering end. They keep on and on researching, but just don't get things done in time." DaimlerChrysler has "paid dearly" for failing to meet research-project development schedules, he adds.

To speed up development and commercialization, the European Commission is about to set up a 'joint technology initiative', which is expected to get approval in November. The management structure of the consortium, which is to include publicly funded labs and leading private companies, is still under discussion. Its main goal, says Loughhead, will be to facilitate public-private partnerships and get robust hydrogen and fuel-cell technologies to the point of commercial take-off by 2015. Engineers — materials, mechanical and electrical — will be needed to study ways of



Joachim Wolf: seeking true figures from car-makers.

"The interesting thing is that the field is still in a state of flux. You can, therefore, be very creative."
— Zdenek Pors

optimizing fuel-cell systems. Electrochemists and materials scientists will tackle fundamental problems. There are also opportunities for colloid chemists, theoretical chemists, experimental physicists and even mathematicians who model the current-voltage curves of a fuel cell. The European Commission is expected to put in between €80 million and €100 million (US\$107 million to US\$134 million) a year.

German advances

"The interesting thing about working on fuel-cell technology is that the field is still in a state of flux," says Zdenek Pors, a mechanical engineer at the institute of fuel cells at Germany's Research Centre Jülich, one of the largest such centres in the world. "You can, therefore, be very creative when it comes to technology development and fixing technical problems." The 29-year-old postdoc left his native Czech Republic in 2002 to work on developing highly concentrated hydrogen gas, from diesel fuel, for use in high-temperature fuel cells. Pors would have liked to go to Paris, but opted for Germany, where fuel-cell research was more advanced.

Germany alone is set to invest more than €1 billion during the next ten years in hydrogen and fuel-cell technologies. The federal ministry of transport will provide some €500 million through a national innovation programme for R&D, and industry is expected to contribute roughly the same amount. The programme could create up to 1,000 jobs for scientists, engineers and technicians, says Detlef Stolten, director of the Jülich institute. Opportunities, in both industry and academia, range from lab technician to institute head. Stolten expects to be recruiting chemists and physicists from as far afield as India and China.

"Hydrogen energy opens the road to carbon-free cars," says Stolten, "and it will no doubt create scores of jobs for skilled scientists."

Quirin Schiermeier is Nature's Germany correspondent.

FUELLING A NEW GOLD RUSH

Current US energy policies suggest that hydrogen technology will be a long-term growth area. "I would not be surprised if jobs tripled in the next few years, given the amount of effort, energy and money being put into the fuel-cell industry," says Jeff Dederer, a product engineering manager with Siemens in Pittsburgh, Pennsylvania.

California has led the way with the development of health- and climate-friendly technologies. In December 2006, the US Environmental Protection Agency approved a waiver, bringing into effect Californian regulations that would require thousands of zero-emission vehicles to come on to the roads in the next decade. Hydrogen fuel cells are a big part of the plan. Vehicles could be ready within a few years, although the

fuel infrastructure will take longer, notes Gary McVay, manager of the fuel-cell programme at the Pacific Northwest National Laboratory in Richland, Washington.

Government-funded labs, private companies and scientists are already collaborating on fuel-cell R&D and manufacture. This was partly spurred by the launch, in 2000, of the Department of Energy's Solid State Energy Conversion Alliance (SECA).

"SECA is still in a research and development state, so there is primarily need for scientific knowledge," says Donald Collins, SECA project manager at the National Energy Technology Laboratory in Pittsburgh.

The focus, he says, is on advancing materials, reducing costs and increasing the power of fuel cells, and there is a growing



Ready to go: a hydrogen fuel station near Munich.

demand for PhD-level scientists and technicians. High-temperature materials scientists are sought after, as are chemical, industrial and plant engineers. "Not enough are trained in this field," says McVay. Specialists in oxide ceramic materials and solid-state chemistry are particularly in demand, he says.

As commercialization nears, there will be more and different job opportunities, says Collins. New skills will also be required in manufacturing and distribution.

"Technicians for production and trouble shooting, but also the marketing and sales support, will then play a much greater role," says Collins.

Carina Lenotti

MOVERS

William Chameides, dean, Nicholas School of the Environment and Earth Sciences, Duke University, Durham, North Carolina



2005–07: Chief scientist, Environmental Defense, Washington DC

2005–present: Regents' professor emeritus, School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia

1998–2005: Smithgall chair in atmospheric sciences, Georgia Institute of Technology, Atlanta, Georgia

When he went to college, Bill Chameides was intent on becoming a lawyer. But he exchanged his legal dreams for science — first physics then atmospheric sciences — and became one of the environment's star witnesses.

It was dumb luck, he says, that led him to first study ozone in the troposphere, which resulted in him developing a theory of tropospheric ozone production while completing his PhD at Yale University. From there he took a postdoc at the University of Michigan, Ann Arbor, working with the current US National Academy of Sciences president Ralph Cicerone, who had just published pioneering work on ozone-depleting chlorofluorocarbons. Chameides found watching Cicerone bridge the science-policy divide inspiring. Although most scientists at the time focused on atmospheric gas chemistry, Chameides laid down the foundation for dynamic atmospheric chemistry between liquids and gases.

Chameides then spent 25 years at the Georgia Institute of Technology, eventually becoming chair of the atmospheric-sciences department where he pursued policy-relevant research. Two years ago, he made an unorthodox move to become chief scientist of Environmental Defense, an advocacy group based in New York City. "I felt I needed to do more to advance the cause of good environmental stewardship," he says.

Chameides thought he was burning his bridges to academia. But he continued to do research, and in doing so kept his career options open. During Chameides' time at Environmental Defense, the group helped to pass landmark laws and won two important Supreme Court decisions. But academia wouldn't let him go. Duke University sought him out to head the Nicholas School of the Environment and Earth Science — a programme noted for its ambitious goals of training students in the physical, biological, social and policy implications of environmental science.

Cicerone says creating a broad programme that merges research and liberal arts agendas is a challenge. But, he adds, Chameides has the qualifications to pull it off. Chameides is already planning to lead the Nicholas School faculty in an assessment of the major environmental and social problems facing the world over the next 25 years, and explore how Duke can best address them. His goal, he says, is to help shape the next generation of scientists, policy-makers and environmental stewards. ■

Virginia Gewin

NETWORKS & SUPPORT

Keep your eye on the goal

The experience of Daniel Rizzuto (see *Nature* **447**, 350; 2007) shows how an adviser's support and a timely decision can lead to a gratifying new start. Rizzuto swapped from studying brain-machine interfaces to working for his former adviser as project manager on a big new contract. But in my experience, such opportunities seldom arise, so determination and sharp job-hunting skills are priceless.

My journey began with a practical objective: to find a research position in the biotechnology industry. But I had no idea how tortuous the road would be. After my PhD studies at the Weizmann Institute of Science in Rehovot, Israel, in the field of drug targeting and delivery in 2000, I thought I'd benefit from a business degree, so I pursued an MBA at an Israeli university.

In my second year I was offered a postdoc at the MD Anderson Cancer Center in Houston, Texas, to study tumour angiogenesis and biomarkers. Although this did not quite fit my career plan, it held the promise of gaining new skills, experiences and contacts. With much enthusiasm, I interrupted my studies and spent the next three years doing lab research. Nevertheless, my end goal was to land a job in biotech, so I was pleased when my mentor established a start-

up based on the lab's research, encouraging trainees to maintain industry-standard lab notebooks, participate in intellectual-property sessions with patent attorneys, apply strict confidentiality policy, meet with investors and prepare presentations.

Then, with my visa about to expire, I thought I'd found what I was looking for on a second postdoc — my new adviser was an internationally recognized scientist with ties to many respected companies. Unfortunately, I soon discovered our outlooks didn't mesh and she didn't share my interest in commercializing science.

Still, I didn't despair. I completed an interview-skills workshop, upgraded my CV, methodically read career sections on the web, got in touch with my network of colleagues in industry and carefully targeted openings. I took a direct and sincere approach with people who might hire me, remaining flexible about job type and confident about my professional value.

Like Rizzuto's, my academic mentors had a role in creating opportunities and independence. The hardships I encountered have definitely helped me navigate the job market and prepare for the future. ■

Limor Chen is a scientist at Teva Pharmaceutical Industries' R&D Initiative in Jerusalem, Israel.

POSTDOC JOURNAL

To teach or not to teach?

One of the biggest surprises of my fledgling academic career is how much I've come to enjoy teaching. Helping students to understand a difficult concept, or seeing their imagination gripped by a new idea, is immensely rewarding. But right now, indulging this passion is not without its dangers. I had a chance to ponder the conflicting risks and rewards during two weeks that I spent out in the field with my Johannesburg students, looking at some of the oldest rocks on the planet.

I enjoyed myself, and because good knowledge of the material is necessary to teach other people properly, I learnt a lot myself about what ancient lavas and sediments — the subjects of our study — could tell us about Earth 3.5 billion years ago.

On the other hand, it could be argued that the short-term nature of my position here means I can ill afford to spend time away from my research. Involving myself in teaching is 'good experience' to put on my CV, but counts for nothing without more tangible qualifications such as a good publication record.

It seems like a bit of a paradox. But to give myself the best chance of getting a teaching position in the future, I may have to ration how much teaching I actually do in the present. ■

Chris Rowan is a postdoctoral student in the geology department at the University of Johannesburg, South Africa.

naturejobs
presents

THE HIGHLIGHT ON CANADA



Helen Wu
Tel: 212-726-9347
Fax: 212-696-9482
Email: h.wu@natureny.com

Lhea Copeland
Tel: 212-726-9261
Fax: 212-696-9482
Email: l.copeland@natureny.com

naturejobs

nature publishing group 



Scientific Director, Robarts Research Institute The University of Western Ontario

Robarts Research Institute, one of Canada's leading medical research institutions, is seeking a Scientific Director who will champion the Institute's success as a centre of world-class science, technological innovation and the commercialization of research. Founded in 1986, in London, Ontario, the Institute's key scientific areas of strength are found in advanced medical imaging, cellular and molecular biology, genomics, immunology, clinical trials and stem cell biology. Working in an interdisciplinary environment, clinical and basic scientists – physicians and physicists, together with biologists and biomedical engineers – investigate some of the most debilitating diseases of our time, from heart disease and stroke to diabetes, Alzheimer's and many forms of cancer. For more information, visit www.robarts.ca.

The University of Western Ontario and Robarts Research Institute have created a new partnership, by fully integrating Robarts as a distinct research institute within Western's Schulich School of Medicine & Dentistry on July 1, 2007. The Institute has undergone significant growth and transformation in the past fifteen years. The current operating budget for the Institute is \$8 million, while the research funding has grown to an impressive \$36 million. There are currently 500 highly dedicated staff and students working in the labs of Robarts' 44 principal investigators.

The Scientific Director will be supported by an Executive Committee, a senior management team (operations, communications, fund and business development) and an Advisory Council on strategic and operational issues. The Scientific Director is responsible for the overall direction, management and operations of the Institute, and will provide vision and scientific leadership throughout the organization.

The ideal candidate will be a Ph.D., M.D./PhD., or M.D., and an outstanding scientist, recognized internationally in a field relevant to the Institute's activities. She/he will possess executive leadership and administrative skills, as well as the ability

to interact effectively with internal, related and external constituencies. As well as an understanding of intellectual property and entrepreneurial business development, of prime importance will be the ability to recognize and promote new and innovative opportunities to enable the Robarts research mission to thrive and grow. The successful candidate will be appointed in a position at the level of Full Professor into the appropriate department(s) in the Schulich School of Medicine & Dentistry.

Consideration of nominations, applications and expressions of interest will begin in the fall of 2007, for an initial 5-year term appointment effective July 1, 2008.

Please respond in confidence to the address shown below. Applicants should have fluent written communication skills in English. All qualified candidates are encouraged to apply; however, Canadians and permanent residents will be given priority. The University of Western Ontario is committed to employment equity and welcomes applications from all qualified women and men, including visible minorities, aboriginal people and persons with disabilities.

Janet Wright & Associates Inc.

174 Bedford Road, Suite 200,
Toronto, Ontario, Canada M5R 2K9
Fax: (416) 923-8311
uwo-robarts@jwasearch.com



Janet Wright & Associates Inc.

Senior-level recruitment for the public and not-for-profit sectors
www.jwasearch.com



Where great minds meet.

Located in Saskatoon, Saskatchewan, on one of Canada's most attractive campuses, the College of Engineering at the University of Saskatchewan is currently offering exciting growth opportunities for graduate students, post-doctoral fellows, and Canada Research Chairs.

Potential candidates will provide leadership in several strategic areas of research including biofuels and bioprocesses, energy and environment, advanced materials and micro/nano devices, transportation and road science, wireless and digital communication, RF MEMS sensor and microsystems technologies, and biomedical engineering. Join a team internationally recognized for teaching, scholarship, research, and innovation.

If multi-disciplinary interactions, advanced facilities like the Canadian Light Source synchrotron, and groundbreaking collaborative research peak your interest, contact the Dean's Office at dean.engineering@usask.ca to find out more.



> www.usask.ca



HIGHLIGHT: CANADA



Change your environment. Find jobs where you'll make a difference

naturejobs

2008 Izaak Walton Killam Postdoctoral Fellowships

Izaak Walton Killam Postdoctoral Fellowships (KPDF) in most fields of study are tenable for up to two years at Dalhousie University, located in Halifax, Nova Scotia, Canada. KPDFs are valued at \$44,000 CDN per year plus travel allowance, a one-time \$3,000 research grant, and a \$1,000 conference travel grant. Applicants must have recently completed a PhD (Jan 2006 or later) at a recognised university and have no current affiliation with Dalhousie University. Application deadline December 15, 2007. Full details available at:

dalgrad.dal.ca/kpdf



DALHOUSIE UNIVERSITY
Inspiring Minds

NW107228R

Visit

www.naturejobs.com

to seriously improve your career prospects.

naturejobs
making science work



"Previewing the future of medicine, the Samuel Lunenfeld Research Institute is driving innovation and excellence in the health sciences."

Dr. Jim Woodgett, Director of Research

- Among the world's premiere biomedical research institutes.
- Competitively funded, driven by excellence, stunning results.
- Leading-edge research powering the knowledge-based economy and the future of medicine.

SAMUEL LUNENFELD RESEARCH INSTITUTE
MOUNT SINAI HOSPITAL
Bright Minds. Big Hearts. The Best Medicine.

NW110396R

600 University Avenue, Toronto, Ontario Canada M5G 1X5 WWW.MSHRI.ON.CA

Advertisement for Position of Director

Institute of Biomaterials and Biomedical Engineering, University of Toronto, CANADA

Nominations and applications are invited for the position of Director, Institute of Biomaterials and Biomedical Engineering (IBBME) at the University of Toronto. The successful candidate should be a dynamic and innovative leader, a world-class researcher with a demonstrated record of excellence in biomedical engineering and qualified for appointment at the rank of Full Professor.

The University of Toronto is the largest and most prestigious research-intensive university in Canada and one of the largest universities in North America with more than 40,000 students registered on its St. George campus and a faculty and staff complement exceeding 11,000. Designated by the United Nations as one of the world's most livable cities, Toronto is demographically diverse and rich in green space, and in cultural and recreational activities.

IBBME is home to leading-edge research and education in biomedical engineering over a broad range of topics from stem cells and tissue engineering to molecular imaging, biomedical nanotechnology and rehabilitation engineering. It is a 'department' of three Faculties (Applied Science and Engineering, Dentistry, and Medicine). With over 30 faculty appointments, including five Canada Research Chairs, 150 graduate students, IBBME is responsible for biomedical engineering graduate and undergraduate programs that rank among the best in the world.

All qualified candidates are encouraged to apply; however, Canadians and Permanent Residents will be given priority. UofT is strongly committed to diversity within its community and especially welcomes applications from visible minority group members, women, Aboriginal persons, persons with disabilities, members of sexual minority groups, and others who may contribute to further diversification of ideas. The University is also responsive to the needs of dual-career couples.

Candidates interested in being considered for this position should submit a brief letter of interest, a curriculum vitae and the names of three references. The search will be open until a suitable candidate is found but priority will be given to applications received by November 15, 2007. Please send applications and nominations to the Search Committee Secretary:

Margaret McKone
Director, Office of the Dean
Faculty of Applied Science and Engineering
University of Toronto
35 St. George Street, Room 170
Toronto, Ontario, Canada M5S 1A4
Phone: (416) 946-8727
Fax: (416) 978-4859
Email: m.mckone@utoronto.ca

NW110583R

Tenure-Track Faculty Position in Experimental Biophysics

Department of Physics, McGill University

We are currently seeking applications for a tenure-track faculty position at the rank of Assistant Professor in the area of Experimental Biophysics, beginning as early as September 2008.

McGill University's Department of Physics offers a rich research environment that includes the availability of advanced probes and sensors, fast imaging and correlation spectroscopies, and the access to atomic manipulation and microfabrication facilities. We are looking for an individual whose vigorous research program will enhance and complement this effort, and we value strong interdisciplinary interactions within and outside our Department. The successful candidate will be a strong teacher and an outstanding researcher. Preferably by mail, and as soon as possible, interested candidates should submit a curriculum vitae, a statement of research goals and plans, a statement of teaching interests and philosophy, and arrange for at least three letters of reference to be sent directly to:

Professor Charles Gale, Chair
Department of Physics
McGill University, 3600 University St
Montreal, Quebec, Canada H3A 2T8

Review of applications will begin January 7th, 2008; applications will be accepted and reviewed until the position is filled. The successful candidate will be supported by a generous start-up package and could be nominated for a Canada Research Chair.

All qualified candidates are encouraged to apply; however Canadian citizens and permanent residents of Canada will be given priority. McGill University is committed to equity in employment.

NW108503R

CLINICIAN SCIENTIST POSITION IN INFLAMMATION RESEARCH

Departments of Medicine and
Microbiology and Immunology,
Schulich School of Medicine & Dentistry

THE UNIVERSITY OF WESTERN ONTARIO

The Departments of Medicine and Microbiology and Immunology are seeking a Clinician Scientist for a faculty position at the rank of Assistant Professor in the general area of Inflammation. Outstanding candidates will be considered at a higher rank. Priority will be given to candidates with research interests and expertise in the areas of infectious diseases, autoimmunity, transplantation, allergy, and viral immunology, which will complement existing areas of research strength in both departments. The successful candidate will hold a clinical academic position in the appropriate Division within the Department of Medicine and will be expected to participate in the teaching programs of the Departments of Medicine and Microbiology and Immunology at both undergraduate and graduate levels.

Candidates must have an MD or equivalent and must be eligible for licensure in the Province of Ontario.

Please see full ad on <http://www.uwo.ca/mni>

Deadline for applications is December

NW110232R

Commencing fall 2007

2 Postdoctoral Positions in Mathematical Modeling

are available at the University of British
Columbia Centre for Disease Control

The successful candidates will begin development of new mathematical and computational tools to study the transmission dynamics of communicable diseases and devise and evaluate strategies to control the spread of disease in populations by providing quantitative analysis to inform policy recommendations and model a) transmission dynamics of (inter-)pandemic influenza; b) transmission dynamics of STIs and blood borne infections; c) hospital outbreaks; d) animal infection transmission.

Candidates should have a PhD in physics, applied mathematics or mathematical biology/epidemiology; be active in theoretical or computational physics, biology/epidemiology, nonlinear analysis and mathematical modeling; have experience with network theory, statistical physics and complex systems.

Appointments are grant funded by CIHR for one year with the potential for extension to 3 years. The incumbents must be independent, adaptable and able to interact and communicate appropriately with colleagues.

Send e-applications including: a cover letter; CV; up to three relevant preprints/reprints and three letters of reference to krista.english@bccdc.ca Informal inquiries by email are welcome.

NW110696R



Dean, Faculty of Science

The University of Windsor invites applications and nominations for the position of Dean, Faculty of Science.

Located in Windsor, Ontario, the University of Windsor has more than 140 undergraduate and graduate programs across nine faculties for 16,000 full- and part-time students. The Faculty of Science has about 135 faculty members and 60 technical and support staff. Departments include Biological Sciences, Chemistry and Biochemistry, Earth and Environmental Sciences, Economics, Mathematics and Statistics, and Physics, and the School of Computer Science, and serve 2,000 undergraduate and 300 graduate students. For additional information, please visit www.windsor.ca/science.

Reporting to the Vice-President Academic and Provost, the Dean of Science will have excellent academic credentials, a strong research record, and proven administrative and leadership ability. The renewable five-year appointment will commence in July 2008.

The University of Windsor is committed to equity in its academic policies, practices, and programs; supports diversity in its

teaching, learning, and work environments; and ensures that applications from members of traditionally marginalized groups are seriously considered under its employment equity policy. Those who would contribute to the further diversification of its faculty and its scholarship include, but are not limited to, women, Aboriginal peoples, persons with disabilities, members of visible minorities, and members of sexual minority groups.

The University of Windsor invites such candidates to apply to its welcoming community and to self-identify in their letter of application. Priority will be given to Canadians and permanent residents of Canada. The Search Committee will begin review of applications and nominations on October 15, 2007. Documentation should be submitted to the address shown below.

Janet Wright & Associates Inc.

174 Bedford Road, Suite 200
Toronto, Ontario, Canada M5R 2K9
Fax: 416-923-8311
uwindsorsc@jwasearch.com

Janet Wright & Associates Inc.

Senior-level recruitment for the public and not-for-profit sectors
www.jwasearch.com



NW110697R



www.careers.ualberta.ca

Assistant or Associate Professor, Microbial Ecology

We invite applications for a tenure-track position at the Assistant or Associate Professor level in research areas related to Microbial Ecology or Environmental Microbiology. We welcome applicants who use modern molecular and/or biogeochemical approaches, and apply them to environmental systems or prokaryote-eukaryote interactions in the natural environment. The successful candidate will have the opportunity to interact with established interdisciplinary groups in microbiology, cell biology, comparative immunology, freshwater and northern ecology, and marine biology. The candidate must have a PhD, two or more years of postdoctoral experience with a strong record of research and have demonstrated potential for excellence in teaching. The University of Alberta offers a competitive salary commensurate with experience and an excellent benefits plan.

The Department of Biological Sciences (<http://www.biology.ualberta.ca/>), with 72 faculty members and 275 graduate students, offers an exciting environment for collaborative research. Exceptional infrastructure includes

molecular biology, microarray and microscopy/imaging services, biogeochemical analysis laboratory and field stations including Meanook and the Bamfield Marine Sciences Centre.

Candidates should submit a curriculum vitae, a one-page summary of research plans, a statement of teaching interests and reprints of their three most significant publications preferably electronically to positions@biology.ualberta.ca

Applicants must also arrange for three letters of reference to be sent to the Chair. The earliest date of employment could be January 1, 2008.

Interested applicants may apply to:

Email: positions@biology.ualberta.ca
Dr. L. S. Frost, Chair
Department of Biological Sciences
CW 405 Biological Sciences Bldg.
University of Alberta
Edmonton, Alberta, Canada T6G 2E9

Deadline: October 15, 2007

All qualified candidates are encouraged to apply; however, Canadians and permanent residents will be given priority. If suitable Canadian citizens and permanent residents cannot be found, other individuals will be considered. The University of Alberta hires on the basis of merit. We are committed to the principle of equity in employment. We welcome diversity and encourage applications from all qualified women and men, including persons with disabilities, members of visible minorities, and Aboriginal persons.

NW109865R



UNIVERSITY OF
ALBERTA
EDMONTON, ALBERTA, CANADA

www.careers.ualberta.ca



Chair, Department of Pharmacology

The Faculty of Medicine and Dentistry at the University of Alberta invite applications for a full-time tenured position as Chair of the Department of Pharmacology. The successful candidate will be an internationally recognized basic or clinical pharmacologist. The individual will provide dynamic and innovative leadership, complement the excellent ongoing research and teaching endeavors within the Department and maintain a strong independently-funded research program.

The Department of Pharmacology has particular strengths in the areas of neuropharmacology of the peripheral and central nervous systems and cardiovascular pharmacology. There are currently 20 academic staff in the Department who occupy research laboratories on the 9th floor of the Medical Sciences Building or in the adjacent Heritage Medical Research Centre. The Alberta Heritage Foundation for Medical Research provides an ideal opportunity to establish new initiatives and attract outstanding investigators. The opportunity also exists to develop new research themes within any of the Research Centres or Groups in the University of Alberta. These include the Centre for Neuroscience, the Mazankowski Alberta Heart Institute, the Alberta Diabetes Research Institute, the Alzheimer's and Neurodegenerative Research Group, the Alberta Asthma Centre, the Cardiovascular Research Group, the Neurochemical Research Unit, the Signal Transduction Group, or the Membrane Transport Group.

The Faculty of Medicine & Dentistry at the University of Alberta is one of the most research-intensive medical schools of Canada, with more than \$160 million in annual research funding. A survey conducted by The Scientist magazine (<http://www.thescientist.com/article/display/14432/>) ranked the University of Alberta among the best places to work in international academia, and in a 2004 survey of 3,500 universities, the University of Alberta was ranked as the best place in North America for post-doctoral fellows. Details about the University of Alberta, the Faculty of Medicine and Dentistry, Edmonton and the Province of Alberta can be found on the Faculty's Home Page at <http://www.med.ualberta.ca/>.

Interested candidates should submit an electronic version of their application outlining their current research interests and leadership experience and their thoughts on how to meet the challenges of academic leadership. The application package should also include an up-to-date curriculum vitae, a short (3-page) statement of current and future research interests and a summary of teaching experience and philosophy. Three reference letters should also be forwarded under separate cover.

Interested individuals are asked to submit their material by October 31, 2007, however, the competition will remain open until the position is filled. Please send your application package to:

Thomas J. Marrie, MD
Dean, Faculty of Medicine and Dentistry
University of Alberta
2J2.00 Walter C. Mackenzie Health Sciences Centre
8440-112 Street
Edmonton AB Canada T6G 2R7
Email: wendy.brown@ualberta.ca

All qualified candidates are encouraged to apply; however, Canadians and permanent residents will be given priority. If suitable Canadian citizens and permanent residents cannot be found, other individuals will be considered. The University of Alberta hires on the basis of merit. We are committed to the principle of equity in employment. We welcome diversity and encourage applications from all qualified women and men, including persons with disabilities, members of visible minorities, and Aboriginal persons.

NW111486R

Shocking Career Prospects?

**Meet better employers
at our regular job fairs.
In the US and beyond.**



naturejobs



HUMAN GENETICIST Tenure-Track/Tenure Position

The newly formed intramural Laboratory of Translational Genomics (LTG) in the Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute (NCI), National Institutes of Health (NIH), Department of Health and Human Services (DHHS), is recruiting two tenure-track/tenured investigators. The mission of the LTG is to investigate the genetic basis of strong association signals identified by candidate gene approaches, linkage analyses in high-risk families, or genome-wide association studies (GWAS), particularly loci identified by the ongoing Cancer Genetic Markers of Susceptibility (CGEMS) program involving GWAS of several major cancers. Investigators in the LTG are expected to develop an independent research portfolio in cancer genomics focused on (1) fine mapping and re-sequencing of loci relevant to cancer susceptibility and/or outcomes, (2) investigation into the causal gene variants that provide biological plausibility for each locus, and (3) bioinformatic analyses of publicly available datasets derived from germline annotation of genetic variation and somatic alterations in cancers. Each investigator is expected to leverage the NCI resources in molecular epidemiology, high-throughput genotyping and whole genome scans, biostatistics and bioinformatics, as well as in basic and clinical sciences. The incumbent will receive research support for developing a state-of-the-art genomics laboratory, and recruiting two post-doctoral fellows/bioinformaticians and a technician.

Applicants must have an M.D. and/or Ph.D. in a relevant field, extensive post-doctoral experience, and a record of publications demonstrating potential for creative independent research in human cancer genetics. Facility with bioinformatics databases and high dimensional data are highly desirable along with strong communication skills. Interested individuals should send a cover letter, curriculum vitae and a brief summary of research accomplishments and goals, along with copies of three to five publications or preprints, and three letters of reference to:

Ms. Judy Schwadron, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Blvd. EPS/8073, Bethesda, MD 20892.

Recommendations can be included with the package or sent directly by the recommender to Ms. Schwadron. Candidates should submit applications by **October 15, 2007**; at this time, the committee will begin to look at suitable candidates. However, the search will continue until qualified scientists are found. Additional information about staff and ongoing research in the NCI Division of Cancer Epidemiology and Genetics is available at <http://www.dceg.cancer.gov>. Please contact **Dr. Stephen Chanock** (phone 301-435-7559 at chanocks@mail.nih.gov) or **Dr. Peggy Tucker** (phone 301-496-8031 at tuckerp@mail.nih.gov) for questions about the position(s).



TENURE-TRACK POSITION CELL BIOLOGY OF HOST-PATHOGEN INTERACTIONS National Institute of Child Health and Human Development

A tenure-track position is available in the Cell Biology and Metabolism Branch (<http://eclipse.nichd.nih.gov/nichd/cbmb/index.html>), NICHD, NIH, to develop an independent research program on the cell biology of host-pathogen or -symbiont interactions. Pathogens and symbionts of interest include viruses, bacteria, and fungi. Outstanding candidates in other areas of cell biology will also be considered. The CBMB has a tradition of excellence in various areas of eukaryotic and prokaryotic cell biology. Other research groups are headed by Irwin Arias, Juan Bonifacio, Ramanujan Hegde, Mary Lilly, Jennifer Lippincott-Schwartz, and Gisela Storz. The recruitment package includes generous funding, two or three additional positions, and laboratory space on the NIH campus in Bethesda. Candidates must have an MD or PhD. Applications should be sent to groverm@mail.nih.gov and include PDF files of the applicant's CV, bibliography, and two-page statement of research plans. Applicants should have three letters of recommendation sent to the above e-mail address. The application deadline is **November 1, 2007**.

Postdoctoral, Research, and Clinical Fellowships at the National Institutes of Health

www.training.nih.gov/pdopenings

www.training.nih.gov/clinopenings

Train at the bench, the bedside, or both

Office of Intramural Training and Education
Bethesda, Maryland 20892
800.445.8283



OPPORTUNITIES @ NIH THE NATIONAL INSTITUTES OF HEALTH

A bullet with your name on it

You have the right to remain silent.

David Hall

Mike, you probably don't recognize my handwriting, what with e-mails, texting and so on, but this is a desperate note scrawled by me, Alan.

First — if you're anywhere near your TIE gun, turn it off. If you're not near it, go get it and turn it off. Take the batteries out. Unload it.

This message has to convince you that you've got to see the Commissioner. Tell him that these new TIE guns have no place in police work. Not in this country. But see him in person. Don't use e-mail or telephones. Nothing electronic is safe.

You remember the armed robbery I was called to a few months back? The incident when I shot that six-year-old kid? The press was full of it. Tragic accident, and so on.

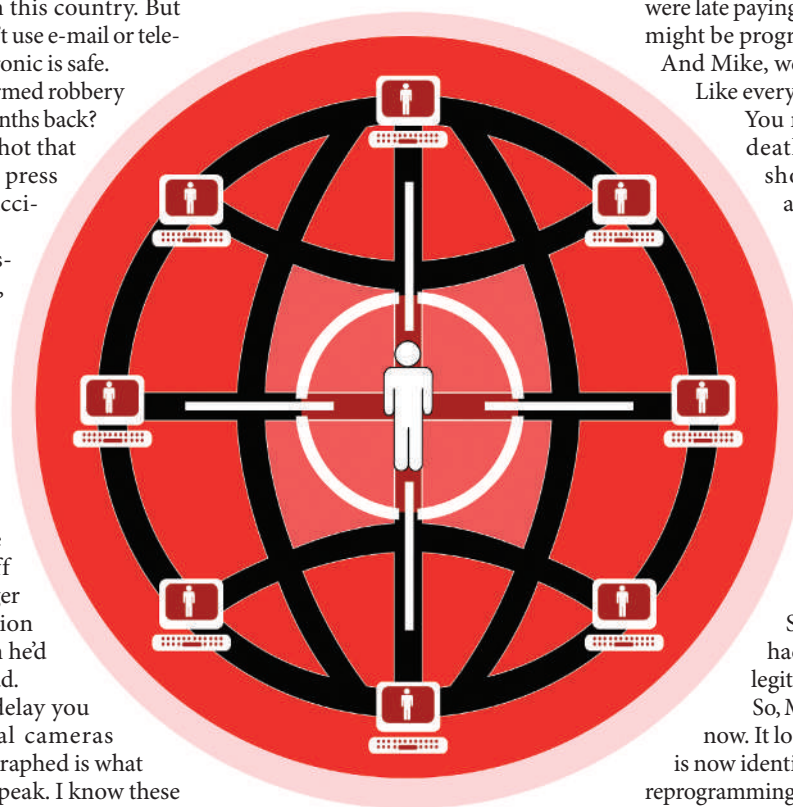
I was glad the investigation cleared me, but most of the details were withheld, and I wasn't sure why. It was my TIE gun that went off and killed the lad. I didn't fire at him. It was the guy with the handgun I was aiming at. But the TIE gun didn't go off when I pulled the trigger — it fired just a fraction of a second later when he'd moved, exposing the lad.

It seemed like the delay you often get with digital cameras — what you've photographed is what happened next, so to speak. I know these new guns have digital cameras attached to the sights.

We don't realize how much electronics have been packed into the new guns since all that fuss in the press about police getting shot with their own weapons when criminals grab them. These new TIE guns are meant to be foolproof, and fire only when the 'right' person pulls the trigger. I thought the fatal delay was something to do with slow response in the identification circuitry.

I'd got a good lawyer onto my defence: a well connected investigator called Julia who could get me the guns' specifications. We thought if we could expose a uselessly slow response time, not only would I be cleared, but we could get the TIE guns banned.

We found something far worse, Mike. Some genius has decided to put another level of identification software in the guns. Something that would let them fire only if the 'right' person was at the business end. Point the gun at an innocent person, and it doesn't go off. How does the gun know who is innocent? Simple — put the equivalent of a cameraphone in each gun to connect it to a national database anywhere there's a wireless link. Use face-identification software to search the national database.



All very clever, but that boy still died. We found that TIE stands for Target Identification and Elimination. The emphasis is on targeting villains, not protecting the innocent.

Julia and I did some more digging. She found out that a DNA profile of the boy had been stored on one national database in a paternity case. A couple of years later it was transferred to another computer to be used in a long-term experiment to see if you could predict criminal behaviour from genetic make-up. The classification system used had identified the boy as a potential mass-murderer. This somehow got into the National Police Computer, and onto the database accessed by the guns. This never

should have happened — the experiment's nowhere near finished yet.

My TIE gun had been told to ignore a real gunman threatening one person (me) and to fire at someone with the potential to kill several in the future.

Julia got me cleared, but she wanted to go public on misuse of personal information held on government computer systems. She said that no one, and she repeated no one, now knew what governed transfer between different national computer systems. She joked that if you were late paying your income tax the guns might be programmed to eliminate you. And Mike, we kept in touch by e-mail.

Like everyone does.

You may have read about her death yesterday in "a tragic shooting accident". It was a TIE gun, but not even being held in a police officer's hand. I went to see the officer. He told me that the gun had started vibrating and he'd dropped it. When it hit the floor, it went off, and Julia happened to be in the way.

I didn't point out that the gun shouldn't have gone off like that. I've just been to check the computer records. Sure enough, Julia's name had been added to the list of legitimate targets.

So, Mike, get these guns banned now. It looks as if the software itself is now identifying people as targets. It's reprogramming the guns to make them fire even if a human hasn't pulled the trigger. The 'legitimate targets' are now not people it's permissible to shoot in the last resort. They are people who will be shot.

If you don't read about my 'tragic death in a shooting accident' it's because I intend to travel where there's no mobile phone or wireless network signal.

As you may have guessed, my name is now on a little list. Of people who won't be missed. Not by those TIE guns.

Mind how you go.

Alan.

David Hall studied biochemistry at Cambridge, worked on yeast long enough to make good bread, then got a proper job, first writing food-science abstracts, then translating technical German.

JACEY